

```

QY 904 INKPEETAKKDRATNSKATTPKOKPTKAPKKKSTKKKTMRYRKRK----- 954
DB 1194 -LNKPAKAEKPTK-----PKKEVSKKEPKTEPPK-APAKKMPEDDDDEPE 1243
QY 955 ---TTPPKKMTSTMP 967
DB 1244 ADFTMPAKKPDTEP 1259

RESULT 11
Q9LIE8
ID Q9LIE8 PRELIMINARY; PRT: 1480 AA.
AC Q9LIE8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SMILIARY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC euroids II: Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP001306; BAB03062.1;
DR Interpro: IPR002965; P_rich_extensions.
DR PRINTS: PR01217; PRICHTENSN.
SQ SEQUENCE 1480 AA; 147153 MW; D1AC0C79F155E732 CRC64;

```

```

Query Match 11.38; Score 800.5; DB 10; Length 1480;
Best Local Similarity 29.68; Pred. No. 1,2e-45;
Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

```

```

QY 116 KKKPKRPVVDGSGGNDNDFKVTTPDSTQHNKVSIS---PKITTKAPINPRSL 171
DB 117 KRPKPKRPVVK-----PHPKRPPTKPHPKRPPTKPHPKRPPTKPKRPV 167
QY 172 PPNSDTSKETSIVNKETVETKTTTNTKQSTGKEKTSKETSISKTSKDLAPT 231
DB 168 KPPSTPK-----PPTTNPPSTPQPPPHKPPPCPTTPPVASBPVATPP 211
QY 232 SKV--LAKPTKATTTGPAITPKKEPT--TTP--KEPASTPKKEPTTTKSAPTTP 285
DB 212 TQMPPIATP-PIAKSPVATPPIATPPIATPPIATPPIATPPIATPPIATPPI 270
QY 286 KEPAPTTT---KSAPTTKEPAPTTT-----KEPAPTTKEPAPTTTKE 326
DB 271 PVAPPTINPISKRPVTP---PTTTPPIAKPPIATPPISTPPATPPAATPPITTP 326
QY 327 PA-----PTTKSAPTPP--KEPAPTTPKKRAP--TTPKEPAPTTPKEPTTPTPKEAP 376
DB 327 PAKPVAISIVT--PVPVPIAQPVATPPTATPVPATPPIATPPISTPSISPS 384
QY 377 TTKEPATTPP--KEPAPAPKRPAPTPPKAPAPTPPKAPAPTTT---KEPSTTPKEAP 431
DB 385 VAPPPATSPKTPPKAPKPPVATPPIAKSPVATPPIATPPIATPPIATPPIATP 444
QY 432 TTKKSAPTTTKEPAPTTTSAPTTPKEP--SPTTKEPAPTTPKEAP---TTPKKPAPT 486
DB 445 TAT---PVAKPPEVETPPIATPPTAPKPISTPPIKSPVATPPIATPPIATP 501

```

```

QY 487 TPKEPAPTTPKPE--APTTTKRAPAP--KEPAPTTPKEAPTT-----PKILTP--T 534
DB 502 TTPPLAIPVAKPPVVTPTATPPIATPPIAKSPVATPPTATPPIATPPIAKPPVVTPT 561
QY 535 TPKEAPTTTPKEPAPTTPPEELAPTT--PEEPPTTPPEEPAPTTPKAAAPNT---PKAPPT 590
DB 562 TPTATPPVAKPPVATPPIATPPIATPPIAKPPVATPPIATPPIATPPIATPPIATP 621
QY 591 TPKEPAPTTPKEPAPTTPKETAP--TTPKGAAPT---LKEPAPTTPKKAPKELAPTT 645
DB 622 TPIATPPIAKPPVATPPTATPPIATSPVAKPPVATPPIATPPIAKPPVATPPIATP 661
QY 646 KEPTSTSDKAPAPTTPKGTAP--TTPKEPAPTTPKEPAPTTPKGAAPTTPKEPAPTTPK 704
DB 682 KPPVAT---PPTATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATP 736
QY 705 APKELAPTTTGPST--TSDKAPAPTTPKETAPT--TPKEPAPTTPKKAPAPTTP-----E 755
DB 737 MPPIATPPTAKPPIATPPIAIPVAKPPVVTPTATPPIATPPIAKSPVATPPTATP 796
QY 756 TTPPTSEVSNPTTPKPTTH--KSPDESTPELSAPTPKALENSPKKEGVTPTTPAA 813
DB 797 TTPIAKPPVATPPTATPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATP 856
QY 814 TKPEMT--TKAKDRTTERDLTTPETTTAAPKMTKETATTTETTESKITATTTQVSTTT 872
DB 857 TPTTTPPAKPPVATPPIATPPI--IAKPPVATPPIATPPIATPPIATPPIATPPIATP 914
QY 873 QDTTPEKITLKTTLAPKATTTTKT---ITTEIMNKPEETAKKDRATNSKATTPK 928
DB 915 PAKPVAIPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATP 971
QY 929 QKPTKAPKKTSTKKTTPKRVKPKTTPPKMTSTPELNP-----TSRTAE--- 977
DB 972 KPEPTTP--PTAP--PVPAMPPIATP---PTAKPVAIPPIATPPIAKPPVATPPIATP 1025
QY 978 -----MLQTTTPNOTPVSKL--VEVNPKSSDAGAGETPHMLLRPHVFEVETPMD 1030
DB 1026 VLPPIAKPPVETSPVATPPTATPPIATPPIATPPIATPPIATPPIATPPIATP 1070
QY 1031 YLPRVNOGIIINPMLS 1047
DB 1071 ATPPVATNPPTAMPPIATP 1087

RESULT 12
Q41805
ID Q41805 PRELIMINARY; PRT: 1188 AA.
AC Q41805;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN PRECURSOR.
OS Zea mays (Maize).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_Taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B73; TISSUE=POLLEN;
RA Rudinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z34465; CA84230.1;
DR Mendel: 14346; Zeama; 2368; 14346.
DR Interpro: IPR001611; LRR.
DR pfam: PF00560; LRR; 3.
DR SMART: SM00370; LRR; 4.
FT SIGNAL.
SQ SEQUENCE 1 27 POTENTIAL.
1188 AA; 120980 MW; 2C77C7F8BD7130149 CRC64;

```

```

QY 664 TAPTP-----KEPAPTTKEPAPTTPK--GTAPTTLKEPAPTTPKKPAKELAPTT 713
DB 994 LPPAPVSSPPVYKSSPPPTPVSSPPPTKPLPPPTPVSSPPPTKXKLPPP---APVS 1050
QY 714 TKGPTSTTSKPAAPT--PKETAPTTPK---EPAPTTKPAAPTTPETPPPTTSVSTP 767
DB 1051 SPPVYKSSPPAPVSLPPTPKRPSPPRTVSSPRVYKCCPPTLVSSPPAPKSLPPP 1110
QY 768 TTTKEPTTIHKSDESTPELSAETPKALENSKREGVPTTKTPATKPEMTTADKTT 827
DB 1111 TPVSSP-----PPEVKSSPPPTPVSSPPAPKSSPPPTPVSSPP----- 1150
QY 828 ERDLRTTP-----ETTAAPKMKETATTEKTESKITATTTQVSTTTQDTTPPKITT 882
DB 1151 --ELKSSPPAPVSSPPSARSSPPAPVSLPPEVKSSPPAPVSSPPAPKSP----- 1203
QY 883 LKTTTLAPKVTYTTTKTITTTTETIMNKPPEETAKPKDRATNSKATTPKPKPAKAKKPTSTK 942
DB 1204 -----PPPA-----PMSSLPPPVKSSPPAPVSSPPAPVSSPPAPKSP 1242
QY 943 KPKTMPRVKAKKTTPTPKKMTSTMPELNPTSRISAEAMLQTTTRPNQTPNSKLVENPKSE 1002
DB 1243 SPPAP--VKPSSLPPAPVSSPPAPV-----TSAP-----PKKE 1275
QY 1003 DAGAEGETPHMLRPVHVEVTPDMDYLPRVNOGIINPMLSDE 1049
DB 1276 EDSTA-----PPAELPPSPFNDIILPPIWANK 1303

RESULT 10
Q20007 PRELIMINARY; PRT: 1274 AA.
AC 020007
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE COSMID F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2; PubMed=7906398;
RX MEDLINE=94150718;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Koopa A., Saunders D., Shownkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RA *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;

```

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U46675; AAB52641.1; C;  
 SQ SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CNGC64;

Query Match 11.4%; Score 802; DB 5; Length 1274;  
 Best local similarity 28.0%; Pred. No. 8.5e-46;  
 Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

QY 112 KNRTRK-KP-----KPVYDEAGSGLDNGD-----FKYTTDOTSTQHNKSTS 156
DB 274 KNPTRKMPWEDETPEVEKPPVPEKKAPVLLKKDAPAKARDDSPSKAAKKEPS 333
QY 157 KRTTAKRP--NRPSLPPNSDTSKETSLSLVNKEETVETKEETTTNNKQTSIDGKEKTTSAK 215
DB 334 SPVPPPTPVKNVYKPKPWEVDEDEPAEVEKPKSAPEKKTIVLKKREPESSSTPSSDPS 393
QY 216 ETQSIKTSKADLAPTSKVLAKPTPKAE-----TTTKGA-----LTTTP 254
DB 394 PKKAAPAVKPRDSSPKKATPLQADEKAEVPTPVKNPVKKYKPPWEVDEDEPAEVEKOP 453
QY 255 KEPTPTP-----KEPASTPKEP-----TP-TIKSAPTTKPEAPTTTSAPTTKPEAP 305
DB 454 EAPAKTIVLKKREPAKADTAFAISKPTETPEKDKDVPKPRDSSPKVAAKPDQAQADA- 512
QY 306 TTKKEPA-----PTTKPEAPTTTKEAPPT-----TTKSAPTTP 339
DB 513 TPVKNVYKMRPWEDETPADSVSKPTDAKKTPLSLAKKDPAKESLKPADTKAPAKP 572
QY 340 KEP-----APTTKPKAPTTPK-----PAPTKEPTPTTKPEAPTTTKEAPTTPK-EP 389
DB 573 RDPSPKVAAPTAEKKTIVLAKKEPAGPADSKTKEPEKSKPRDPSKKAAPVAPKTEV 632
QY 390 APTAPKKAP-----TTKPEAPTTTKEAPTTTKEP-----SPTTKPEAPTTTKSAPT 439
DB 633 APAVKKPELISKPKDAPAKKAEPSVVP--PTPVKNPVKKYKPPWEDEDAKPVSLDE 691
QY 440 TTKPEAPTTTKSAPTTT-KEPSPTTKKEPAPTTPK--EPAPTTPKKAP--TTKPEAPT 494
DB 692 PEKK--TPVLAKKAPTAKDSEADPAVSGPSSKPLAKKAPKPRDPSPKAAVPIKAPK 750
QY 495 TPKPEAPTTTTPKAPTA--KPEAPTTTKEPAPTTTTPKLTPTP-----EKLAPTTEKPA 548
DB 751 T--EVPVAVYKKEPAPVAKSRDPSPKAK--AEPNSP--VVPPTPVKNPVKKYKPPWEDDA 805
QY 549 PTPEELAPTTPEP-----TPTPEEPAPTTTAKAAPTTPKPEAPTTTKEAPTTT-PE 602
DB 806 PAEPPVNPPEPEKKPPVLAKKTPVAKPRDPSKAPVAPASTKTADAPVSVKPPPEVSKPE 865
QY 603 PAPTTPKETAPTTKGAPTTTLLKEPAPTTTTPKPAKELAPT--TTKEPTSTTSOKPAPTTP 661
DB 866 PSPKKAEPNSVVP-----PTPVKNPVKKW--KPWEDEDEPTTEVKKPSE--PEKKTIVLA 918
QY 662 KGTAPTTKPEAPTTTKEPAPTTTTPKGAPTTLLKEPAPT-----TTKKAP-----KEL 709
DB 919 K-KEPEKPKD-APKVAAPRDPSPKAPVPE--KPEKAVAKPRDLSKKAIPIPANQEA 974
QY 710 APTTTKGPS-----TSDKP-----APTTPKET-----APTTPEAPTTTTPKAPT 752
DB 975 PPTPVKNPVKKYKPPWEDEDEPAEVSAPPEKKTIVLAKKAPAKPRDP--SPKKAAPV 1031
QY 753 PPTPEPTTSEVSTPTTKEPTTIHKS-----DESTPELSA-EP-----TPKALENSPKEPG 804
DB 1032 AAK--PDPKIPV--PPTPVKNPVKKYKPPWEDEDESEVSAPEPKKTIVLAKKAPTKPA 1089
QY 805 V-----PPTTKTPAAT-----KPEMTTAKDKTTERDLRTTPEPT--TAAPK 843
DB 1090 TKPDSEAADPVSGPTSKPKLSKAPVKEKPTTDPKDDKTKPSPAKPKPEAPAPAPK 1149
QY 844 MKKETATTEKTESKITATTTQVSTTTQDTTPPKITTLTKTTIAPVTTTKTITTE 903
DB 1150 KKKPVMDDDPDEEADFTVPAPSKKPDTEADPDLG-----GPKTKDPK----- 1193

```

QY 227 DLAPTSVLAKPPTKAPETTTGSPALTTPKRPPTPKPAPSTPKPPTTKSAPTTRK 286  
 DB 478 TYAPTKETVAPT---EETTYASTEEYVAPTEETTYAPAEETPYEETET--TYAPTEET 533  
 QY 287 EPAPT--TKKAPPTPKPAP--TKKAPPTPKPAPPTTKKAPPTTKSAPT----- 337  
 DB 534 TYAPTEETTYAPTEETTYAPAEETPYEETETTYAPTEET--YAPTEETMYA 592  
 QY 338 ---TKKAPPTPKKAPPTPKPAPPTTKPPTPKP-----AP--- 376  
 DB 593 PTEETTYAPTEETTYAPAEETPYEETETTYAPTEETTYAPTEETTYASTEEYVAPTE 652  
 QY 377 TTKKAPPTPKP-----APAKKAPPTPKKAPPTTKKAPPT--TKKPSPTPK 428  
 DB 653 TYAPAEETPYEETETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 712  
 QY 429 P-----APT--TKKAPPT-----TKKAP-----APT--TKKAPPT--TKKPSPT 462  
 DB 713 PTEETTYAPTEETTYAPTEETMYAPTEETTYGPTTEETTYAPTEATVYAPTEETPYAPTE 772  
 QY 463 TTKP-----APTPKAPPTPKKAPPTTKKAPPT--TKKAPPTTKKAPPTAK 514  
 DB 773 TYEPGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 832  
 QY 515 PAPTPKETAPPTPKKAPPTTKKAPPTPKKAPPTPKKAPPT--TKKAPPTPKP 571  
 DB 833 PTEET--TYTPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 890  
 QY 572 ----APT-----TPKAAAPTPKKAPPTPKP-----APTTPKAPPTPK 610  
 DB 891 KETTYAPTEETTYASTEEYVAPTEETTYAPAEETPYEETETTYAPTEETTYAPTE 950  
 QY 611 TAPTTPKGTAPT--TKKAPPTPKKAPK--ELAPT--TKKPSPTTKKAPPTPKGTA 665  
 DB 951 YAPTEETTYAPTEETTYAPAEETPYEETETTYAPTEETTYAPTEETTYAPTE 1010  
 QY 666 PTPKAPPTPKP-----APTTPKGTAPT-----TKKAPPTPK 701  
 DB 1011 EETTYAPAEETPYEETETTYAPTEETTYAPTEETTYASTEEYVAPTEETTYAPAE 1070  
 QY 702 KKP-----AKP-----LAPT--TKKPSPTTKKAPPTPKKAPPTPKKAPPT 750  
 DB 1071 YEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEETETTYAP 1130  
 QY 751 PTPTEPTPKPSTSEVSTPTTKPPT--TKKPSPTTKKAPPTPKKAPPTPKKAPPT 804  
 DB 1131 EETTYAPTEETMYAPTEETTYGPTTEETTY--APTEATVYAPTEETPYAPTEETTYEPT 1189  
 QY 805 --VPTTPKAPPTPKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT 1190  
 DB 1190 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1244  
 QY 859 KITATTOVSTTQOQT-----PKTIT--LKTTTAPKVTTKKAPPTTKKAPPT 905  
 DB 1245 TTYAPTEETTYAPTEETTYAPTEETMYAPTEETTYGPTTEETTYAPTEATVYAPTE 1304  
 QY 906 NKPEETAKPKDRAT--NSKATPKP--OKPTKAPKPT-----STKAP--KTMP 949  
 DB 1305 PTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1364  
 QY 950 VRRKPTPTPKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT 1365  
 DB 1365 TDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDEPTDE 1420

GN PEX2.  
 OS Zea mays (Maize).  
 OC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OK NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=POLLEN.  
 RA Stratford S., Barnes W., Golbiewski A., Colter R., McCormick S.,  
 RA Hohorst D., Gao M., Showalter A., Bedinger P.A.;  
 RT "Polien Extensin-Like (Pex) Genes in a Monocot and a Dicot."  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF159297; AAD55980.1;  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR001998; Xylose\_Isom.  
 DR InterPro: IPR002965; P\_Rich\_extensn.  
 DR InterPro: IPR003592; LRR\_out.  
 DR Pfam: PF00560; LRR; 3.  
 DR PRINTS: PR01217; PRICEXTENS.  
 DR PROSITE: PS00172; XYLOSE\_ISOMERASE\_1; UNKNOWN\_1.  
 DR SMART: SM00370; LRR; 3.  
 SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 11.5%; Score 811.5; DB 10; Length 1315;  
 Best Local Similarity 24.6%; Pred. No. 2e-46;  
 Matches 262; Conservative 89; Mismatches 473; Indels 243; Gaps 35;

QY 73 RCFSEFGRGCCDQACCKYDCCEPESFCALVNDKKNRTKKKPTP-----K 122  
 DB 390 QCAPVLAARPEVNCCKH-----VCAGY-----PRGCGPSSSVGK 425  
 QY 123 PLYVDEAGSLDNGDKRVTPDSTTQHNVSTPKITAKPINKPSLP----- 173  
 DB 426 PPSVPG-----KPAAPAMPPTPIPPDVSP-----EPL-PEPSVVPAPAKRMPT 469  
 QY 174 -NSDTSKETSLEYNKETETETETTTNKOTSTDGKEKTSKAKETOSIKTSAKDLAPTS 232  
 DB 470 LRSPAPADEVITP-----PPVPAKSPGTSPPASGAPLQAPPAASSPPATPVKSSPPA 525  
 QY 233 KV-----LAKPTKAEETTTGSPALTTPKRPPTPKPAPSTPKPPTTKSAPT 278  
 DB 526 AVLLPPKAPTPSPAPVAPSPPEAPVSSPQOVKSSPPAPVAPVSSPPPKSP--PPAPV 583  
 QY 279 KSAPTTPKAPPTTKSAPTTP--KEPAP-----TTKAPPTPKKAPPTTKKAPPT 330  
 DB 584 ASPPPLMKSPPPAPVAPSPQPLKSPPEVTLSTPSVKSPPPPVAPVAPSPPPVAPSP 643  
 QY 331 TTKSAPPTPKP-----PAP--TTP-----KKAPPTPKKAPPTPKP--PTTPPKKAPPTPK 379  
 DB 644 AVSSSPSPVKKLPPLAPAKSTTPPEEEKPVPPTPVKSSPPEKSLPPTTLTSPPOEK 703  
 QY 380 EPAPTTPKAPPTAKKAPPTPKKAPPTPKKAPPTPKKAPPTTKKAPPTTKKAPPT 439  
 DB 704 PTPSPKSPPPSPPEVETLPPSKSSPPEEPVSSPQAPKSSP-----PAPVS--SPP 756  
 QY 440 TTKKAPPTTKSAPTTPKPSPTTKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPT 499  
 DB 757 LKSSPPVPESSPPTPKSSPPLAVSSPPQVETKSPAPVSSPPTPKSSPPLAVSS 816  
 QY 500 APTTKKAPPTAKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPT 559  
 DB 817 PPQVEKTSPPAPVSSPPTPKSSPPLAVSSPPQVETKSPAPVSSPPLKPPSPSS 876  
 QY 560 PPEPTTPTEPPAPPTPKAAAPNTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPT 619  
 DB 877 SVSSPPTVKSPPAPVSSPPTPKSSSPPAHVSSPPEAKSSPPLAPVSSPSSPSS 936  
 QY 620 A-----PTTKKAPPTTKKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 663  
 DB 937 SSPPVPEKTSPPATVSSPPTPKSSPP-----AVSSPPVAVSSPPAPVSSPPTPKP 993

RESULT 9  
 Q9SPMO PRELIMINARY; PRT; 1315 AA.  
 AC Q9SPMO;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE EXTENSIN-LIKE PROTEIN.







DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE Y51B11A.1 PROTEIN.  
 GN Y51B11A.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Poloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."  
 RT Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Cotton M.;  
 RT "The sequence of C. elegans cosmid Y51B11A."  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC006797; AAF60743.1;  
 DR InterPro; IPR002965; P-rich\_extensn.  
 DR PRINTS; P01217; PRICHEXTENS.  
 SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 13.8%; Score 971.5; DB 5; Length 1079;

Best Local Similarity 29.5%; Pred. No. 2.9e-57;  
 Matches 330; Conservative 129; Mismatches 481; Indels 177; Gaps 44;

QY 106 EVKDNKNNKTKKPPKPPVVDAGSGLDNGDFKV-----TPDDSTTQ-HNNVSTSP 157  
 DB 31 ETADCRANQTPHTMLP-----STLSVDMETPSLVLSTPSSSTPIKETTTTAP 81  
 QY 158 KITAKPIMPRLPSNDSKETSIVNKEET-VEKETEYTNKQSTDGKEKTTSAKE 216  
 DB 82 ETTSTEP-----PSSSTTPVQTTTITAPETTSREAPSSSTTPVQTT-----TTTAP 128  
 QY 217 TOSIEKTSKADLAPTSKVLAKPTPKAET-TPGP-ALTPPKPEPTTPPKPEASTPKPEPT 274  
 DB 129 TITSTEPSSS---TSPVQTTTITAPETTSREAPSSSTTPVQTTTITAPETTSSTEPSSS 184  
 QY 275 PTIKSAPPTPKPAPTTTKSAPTPKPAPTTKPEAPPTTKPEAPPTTKPEAPPTTKS 334  
 DB 185 TSPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 240  
 QY 335 APPTPKPEAPPTPKPAPTTKSAPTPKPAPTTKPEAPPTTKPEAPPTTKPEAPPTPK 394  
 DB 241 APETTSSTEPSSSTTPVQTTTITAPETTSREAPSSSTTPVQTTTITAPETTSSTEPSSS 299  
 QY 395 KKPAPTPKPEAPPTPKPE-----APPTTKPSPTTKPEAPPTTKKAPPT 440  
 DB 300 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 359  
 QY 441 TK--EPAPPTTKSAPTPKPEPTTKPE-APPTPKPEAPPTTKKPAPTTKPEAPPTPK 497  
 DB 360 TREPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSST 415  
 QY 498 EPAPPTTKKPAPTPKPE-APPTPKETAATPKKLLPTTPPEKIAPTPKPEAPPTPEEL 555  
 DB 416 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 471  
 QY 556 APPTPEPTPTTPPEAPPTPKKAAPTPKPE-APPTPKPEAPPTTKPEAPPTPKETAP 613  
 DB 472 PETTSTEP-PSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSS 529  
 QY 614 TTPKGTATPTTKPEAPPTPKKPAPEKELAPTTTKEPTSTSDKP-APPTPKGTATPTTKPE 672

DB 530 TTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAP 588  
 QY 673 APPTPKPE-APPTPKGTATPTTKPEAPPTPKKPAPEKELAPTTTKEPTSTSDKP-APT 729  
 DB 569 --TITSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSST 646  
 QY 730 PKETAPPTPKPEAPPTPKK-----PAPTPETPPPTTS-----EVSPTTTPKEPTT 776  
 DB 647 PVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETS 706  
 QY 777 HKSPDESPELSAPPT--PKALENSPKPEGVPTTKPAATPEMTTAKDKTERDLRT 834  
 DB 707 TEPPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTT 766  
 QY 835 ---PETTAAPKMKETAATTEKTESKITAATTOVSTTQODTPPKITLKTITLAP 890  
 DB 767 TTTAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTTTITAPETTSSTEP 824  
 QY 891 KYTTTKKTTTTEINMKPPEETA-KPKDRATNSKATTPKPKPKPAKPKPTSKPKTPMR 949  
 DB 825 PSSSNTPVQTTT--TTAPETTSSTEPSSSTTPVQTTTITAPETTSSTEPSSSTTPVQTT 882  
 QY 950 VRKPKTTPPKKMTSPMELNPTSRIAEMLQTTTRPNQTPNSKIVENPKEDGAGAG 1009  
 DB 883 ITAPETTSSTEPSSSTTPVQTTTITAPET--TSTEP--PSSSTTPVQTTTITAPETTS 936  
 QY 1010 -ETPHMLLRHVMPETP---DMOYLPRV-----PNOGIINPMLSDEFN---- 1051  
 DB 937 TEPPSSSTTPVQTTTITPCDCSLSIDRVYPTTEEMWENKRDIIQSYDSPRTAFESF 996  
 QY 1052 -----ICNGKPVGDGLTLRNGTLVAFRGHYWMLS-----PSPSPPAR 1091  
 DB 997 VSTPDICTATLILCIATSGGISNL-NATL-----FLGSDGSSIDLPE----- 1040  
 QY 1092 ITTEWGLPSPIDVYTRCNCEGKTFPEFKSQYKFTN 1128  
 DB 1041 ---YNVPGLG-TPMEINCEGKN-----WSYNN 1064

RESULT 6  
 ID 0917S1 PRELIMINARY; PRT; 1049 AA.  
 AC 0917S1;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE CG5228 PROTEIN.  
 GN CG5228.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abirl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borliva D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,

[illegible]

```

QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 147
Db 181 KIKSSKSNANRELQKLLKVNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 240
QY 148 TOHNKVSSTPKITTAAPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 207
Db 241 TOHNKVSSTPKITTAAPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 300
QY 208 KEKTSKAKETOSIEKTSADLAFTSKVLAKPPKAETTTKGALTTPKKEPTTPKEPAS 267
Db 301 KEKTSKAKETOSIEKTSADLAFTSKVLAKPPKAETTTKGALTTPKKEPTTPKEPAS 360
QY 268 TTPKEPTTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP 327
Db 361 TTPKEPTTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP 420
QY 328 APPTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 387
Db 421 APPTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 480
QY 388 EPAPTTAKKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 447
Db 481 EPAPTTAKKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 540
QY 448 TTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 507
Db 541 TTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 600
QY 508 APAPKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 567
Db 601 APPTTKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 660
QY 568 PEAPAPTTPKAAAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 627
Db 661 PEAPAPTTPKAAAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP 720
QY 628 APPTTKKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 687
Db 721 APPTTKKBPAPKELAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 780
QY 688 TAPPTLKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAP 747
Db 781 TAPPTLKBPAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAP 840
QY 748 KAPAPTPPPPTTSVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 807
Db 841 KAPAPTPPPPTTSVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTT 900
QY 808 TKTPAATKPEMTTAAKOTTERDLKTTTETTTAAKMKKETATTTTEKTESKITATTTQV 867
Db 901 TKTPAATKPEMTTAAKOTTERDLKTTTETTTAAKMKKETATTTTEKTESKITATTTQV 960
QY 961 TKTPTODTTPFKITTLKTTLAPKVTTLTKKLTITTEINMKPEELAPKDRATNSKATPK 927
Db 961 TKTPTODTTPFKITTLKTTLAPKVTTLTKKLTITTEINMKPEELAPKDRATNSKATPK 1020
QY 928 POKPTKAPKPKPTSTKPKKPTMPKPKPTTPPKPKATSTYMBELNPNRIEAMQOTTRN 987
Db 1021 POKPTKAPKPKPTSTKPKKPTMPKPKPTTPPKPKATSTYMBELNPNRIEAMQOTTRN 1080
QY 988 QTPRSKIVENPSESDAGAGETPHMLLRPHFMPEVTPDDMDYLRVPRNOGIINPMLS 1047
Db 1081 QTPRSKIVENPSESDAGAGETPHMLLRPHFMPEVTPDDMDYLRVPRNOGIINPMLS 1140
QY 1048 DETNIGCKPVDGLTTLRNGTILVAFRGHYTWMLSPPSPSPARITFVWGISPIDVFT 1107
Db 1141 DETNIGCKPVDGLTTLRNGTILVAFRGHYTWMLSPPSPSPARITFVWGISPIDVFT 1200
QY 1108 RCNCEGKTFPFKDSQWYRFTNDIKDAGYRPPIFGFGGLGQIYALALSTAKYNNWESVY 1167
Db 1201 RCNCEGKTFPFKDSQWYRFTNDIKDAGYRPPIFGFGGLGQIYALALSTAKYNNWESVY 1260
QY 1168 FFKRGSIOQYIYKQEPVQKCGRRPALNYPVGEEMTVRRRRRPERAIGPSQHTIRIQY 1227

```

```

Db 1261 FFKRGSIOQYIYKQEPVQKCGRRPALNYPVGEEMTVRRRRRPERAIGPSQHTIRIQY 1320
QY 1228 SPARLATODKGVLNENEKVSILMHGILPNVVTSAISLPNIFKPPGDYDYAFESKQYNNIDV 1287
Db 1321 SPARLATODKGVLNENEKVSILMHGILPNVVTSAISLPNIFKPPGDYDYAFESKQYNNIDV 1380
QY 1288 PSRTARATITRSQOTLSKWYNCP 1311
Db 1381 PSRTARATITRSQOTLSKWYNCP 1404

RESULT 3
ID QJUM99 PRELIMINARY; PRT: 1054 AA.
AC QJUM99;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRC4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034730; BAA92310.1;
DR MGD: MGI:1891344; Prg4.
DR InterPro: IPR000365; Hemopexin.
DR InterPro: IPR002965; P_rich_extensin.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PR01217; PRICHEXTENSIN.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match 51.5%; Score 3640.5; DB 11; Length 1054;
Best Local Similarity 52.4%; Pred. No. 1,2e-236;
Matches 740; Conservative 55; Mismatches 159; Indels 457; Gaps 22;

QY 1 MAWTLPIYLLLSVENVIOVSSODLSSCAGRCGEGYSRDATCNDYKCOHMECCPP 60
Db 1 MGWTLIPGCSLLLPVLLIOVSSODLSSCAGRCGEGYSRDATCNDYKCOHMECCPP 60
QY 61 KRVTAEISCGRCGFESREERRECDCAOCKKXKCPDYESFCAB----- 106
Db 61 KRVTAEISCGRCGFESREERRECDCAOCKKXKCPDYESFCAB----- 106
QY 61 KRVSPELSCGRGCFESFARRECDOSQCKQYKCCADIDSFEELVHNSTSPSKATP 120
Db 61 KRVSPELSCGRGCFESFARRECDOSQCKQYKCCADIDSFEELVHNSTSPSKATP 120
QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 180
Db 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 180
QY 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 180
Db 107 -----VKDNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 180
QY 181 SKSNANRELQKLLKVNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 240
Db 181 SKSNANRELQKLLKVNKKNRKTKKPKPPVVDGAGSLDNGDKVTTPST 240
QY 151 NKVSTSPKTTAKKPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 210
Db 151 NKVSTSPKTTAKKPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 210
QY 241 SKVATSPKTTAKKPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 299
Db 241 SKVATSPKTTAKKPINPRSLPNSDTSKETSLTVNKEETTVETKETTNNKQTSIDG 299
QY 211 TTSKAKETOSIEKTSADLAFTSKVLAKPPKAETTTKGALTTPKKEPTTPKEPAS 270

```

DR PRINTS; PRO0022; SOMATOMEDINB.  
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN\_1.  
 DR PROSITE; PS00524; SOMATOMEDIN\_B; 2.  
 DR SMART; SM00120; HX; 2.  
 DR SMART; SM00201; SO; 2.  
 SO SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 99.2%; Score 7007.5; DB 4; Length 1404;  
 Best Local Similarity 93.4%; Pred. No. 0;  
 Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

1 MAMKLPYIYLLLLSVFVIOOVSSODLSSCAGRCGEGYSDATNCNDYNCQHMECCPDF 60  
 1 MAMKLPYIYLLLLSVFVIOOVSSODLSSCAGRCGEGYSDATNCNDYNCQHMECCPDF 60  
 61 KRVCTAELSCGRCFESEREGRECDCAOCKYDKCCPDYESFAE----- 106  
 61 KRVCTAELSCGRCFESEREGRECDCAOCKYDKCCPDYESFAE----- 106  
 107 ----- 106  
 121 PPSGASQIKSTTKRSRPPNKKTKKVIKYESEITEEHSVENQESSSSSSSSSSSTIW 180  
 107 -----VKDNKKNTKKKPTPKPVYDEAGSGLDNGDFKVTTPDST 147  
 181 KIKSSKSAANRELQKLVKDKNKKNTKKKPTPKPVYDEAGSGLDNGDFKVTTPDST 240  
 148 TQHNKYSTPKITTAQINRPSLPNSDTSKETSJLVNKETTVEKETTNNKOTSTDG 207  
 241 TQHNKYSTPKITTAQINRPSLPNSDTSKETSJLVNKETTVEKETTNNKOTSTDG 300  
 208 KEKTSKAKETQSIKTSANDLAPTSTVLAKEPTKKAETTKGPAITTPKEPTTPKEBAS 267  
 301 KEKTSKAKETQSIKTSANDLAPTSTVLAKEPTKKAETTKGPAITTPKEPTTPKEBAS 360  
 268 TTPKEPTTPKISAPTTKEPAATTTKSAPTTPKKEPAATTTKEPAATTTKEP 327  
 361 TTPKEPTTPKISAPTTKEPAATTTKSAPTTPKKEPAATTTKEPAATTTKEP 420  
 328 APTTTSAPTTKEPAATTTKRAATTPKKEPAATTTKEPTTPKEPAATTTKEP 387  
 421 APTTTSAPTTKEPAATTTKRAATTPKKEPAATTTKEPTTPKEPAATTTKEP 480  
 388 EPAPTPKRAATTPKKEPAATTTKESPTTPKKEPAATTTKSAPTTPKEPAAT 447  
 481 EPAPTPKRAATTPKKEPAATTTKESPTTPKKEPAATTTKSAPTTPKEPAAT 540  
 448 TTKSAPTTKESPTTPKKEPAATTTKRAATTPKRAATTPKKEPAATTTK 507  
 541 TTKSAPTTKESPTTPKKEPAATTTKRAATTPKRAATTPKKEPAATTTK 600  
 508 APTAKKEPAATTPKEZAPTTKILPTTPPEKLAATTPPEKAPATTPPELAATTP 567  
 601 APTAKKEPAATTPKEZAPTTKILPTTPPEKLAATTPPEKAPATTPPELAATTP 660  
 568 PEEPAATTPKAAANPTKEPAATTTKRAATTPKKEPAATTTKGAATTPK 627  
 661 PEEPAATTPKAAANPTKEPAATTTKRAATTPKKEPAATTTKGAATTPK 720  
 628 APTTPKRAPELAPTTKEPTSTSDKAPATTPKGAATTPKEPAATTPK 687  
 721 APTTPKRAPELAPTTKEPTSTSDKAPATTPKGAATTPKEPAATTPK 780  
 688 TAPTTLKEPAATTPKRAPELAPTTKGTSTSDKAPATTPKEAATTPKEPAATTPK 747  
 781 TAPTTLKEPAATTPKRAPELAPTTKGTSTSDKAPATTPKEAATTPKEPAATTPK 840  
 748 KPAATTPERTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPKALENSPKRGVPT 807  
 841 KPAATTPERTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPKALENSPKRGVPT 900  
 808 TKTPAATKPEMTTAAKDTTERDLRTTPETTTAAKMKKETATTTKESKITATTTTGV 867

Db 901 TKTPAATKPEMTTAAKDTTERDLRTTPETTTAAKMKKETATTTKESKITATTTTGV 960  
 QY 868 TSTTQDTPPFKITTLKTTTLAPKYTTTKKTTTTTEINKEPEFAKPKDRATNSKATTPK 927  
 Db 961 TSTTQDTPPFKITTLKTTTLAPKYTTTKKTTTTTEINKEPEFAKPKDRATNSKATTPK 1020  
 QY 928 PCKPTKAKKPKPTSTYKPKTMPRVKPKTTPPRKMTSTMPELNPTSRIAEMIQTTTRPN 987  
 Db 1021 PCKPTKAKKPKPTSTYKPKTMPRVKPKTTPPRKMTSTMPELNPTSRIAEMIQTTTRPN 1080  
 QY 988 QTPNSKLEVPNPKSEPDAGAGETPHMLLRHVFMPEVTDMDYLPRVNOGIINPMLS 1047  
 Db 1081 QTPNSKLEVPNPKSEPDAGAGETPHMLLRHVFMPEVTDMDYLPRVNOGIINPMLS 1140  
 QY 1048 DETNLCNKKPYDGLTTLRNGTLVAFRHYFMMLSPFSPSPARRITVWGISPIDVTFT 1107  
 Db 1141 DETNLCNKKPYDGLTTLRNGTLVAFRHYFMMLSPFSPSPARRITVWGISPIDVTFT 1200  
 QY 1108 RCNCEGKTFEFDKSOYRFTNDIKDAGYPRIFKGFGLTGQIVAAIATAKYNMPESEY 1167  
 Db 1201 RCNCEGKTFEFDKSOYRFTNDIKDAGYPRIFKGFGLTGQIVAAIATAKYNMPESEY 1260  
 QY 1168 FFKRGSSIQOYIYKQEPVOKCPGRPALNYPVYGEVTVRRRRFERAIGSPQTHIRIQY 1227  
 Db 1261 FFKRGSSIQOYIYKQEPVOKCPGRPALNYPVYGEVTVRRRRFERAIGSPQTHIRIQY 1320  
 QY 1228 SPARLAYQDKGVLAHNEKVSILMRGLPNVTSAISLPNIRKRDGDYVAFSKDOYNTIDV 1287  
 Db 1321 SPARLAYQDKGVLAHNEKVSILMRGLPNVTSAISLPNIRKRDGDYVAFSKDOYNTIDV 1380  
 QY 1288 PSTTARAITTRSGQITLSKYWNCP 1311  
 Db 1381 PSTTARAITTRSGQITLSKYWNCP 1404

RESULT 2  
 ID Q9BX49 PRELIMINARY: PRT: 1404 AA.  
 AC Q9BX49:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE BG14L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).  
 GN BG14L6.2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wray P.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL133553; CAC36090.1;  
 SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;

Query Match 99.0%; Score 6992.5; DB 4; Length 1404;  
 Best Local Similarity 93.2%; Pred. No. 0;  
 Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;

1 MAMKLPYIYLLLLSVFVIOOVSSODLSSCAGRCGEGYSDATNCNDYNCQHMECCPDF 60  
 1 MAMKLPYIYLLLLSVFVIOOVSSODLSSCAGRCGEGYSDATNCNDYNCQHMECCPDF 60  
 61 KRVCTAELSCGRCFESEREGRECDCAOCKYDKCCPDYESFAE----- 106  
 61 KRVCTAELSCGRCFESEREGRECDCAOCKYDKCCPDYESFAE----- 106  
 107 ----- 106  
 121 PPSGASQIKSTTKRSRPPNKKTKKVIKYESEITEEHSVENQESSSSSSSSSSSTIR 180

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:30:16 ; Search time 210.7 Seconds  
(without alignments)  
910.124 Million cell updates/sec

Title: AAB  
7064  
Sequence: 1 MAMKTLPIYLILLISLVFVIO.....ARAITTSGQTLKRWKNCNP 1311

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPTREMBL.17:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.rodent:\*  
13: sp.virus:\*  
14: sp.vertibrate:\*  
15: sp.vertebrate:\*  
16: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7007.5	99.2	1404	4 Q92954	Q92954 homo sapien
2	6992.5	99.0	1404	4 Q9BX49	Q9BX49 homo sapien
3	3640.5	51.5	1054	11 Q9JN99	Q9JN99 mus musculu
4	1713.5	24.3	401	6 Q77765	Q77765 bos taurus
5	971.5	13.8	1079	5 Q9N457	Q9N457 caenorhabd
6	936.5	13.3	1049	5 Q9I761	Q9I761 drosophila
7	865.5	12.3	1795	5 Q9I649	Q9I649 phytothor
8	855	12.1	1489	10 Q9SPM0	Q9SPM0 zea mays (m
9	811.5	11.5	1315	10 Q9SPM0	Q9SPM0 zea mays (m
10	802	11.4	1274	5 Q20007	Q20007 caenorhabd
11	800.5	11.3	1480	10 Q9LIF8	Q9LIF8 arabidopsis
12	784	11.1	1188	10 Q41805	Q41805 zea mays (m
13	776.5	11.0	2187	11 P70670	P70670 mus musculu
14	752	10.6	2112	5 Q9VEL9	Q9VEL9 mycobacteri
15	745	10.1	763	2 Q9VHL2	Q9VHL2 chlamydomon
16	707.5	10.0	555	2 Q9EP06	Q9EP06 gallus gall
17	675.5	9.6	2284	5 Q9VPG1	Q9VPG1 gallus gall
18	671	9.5	1151	11 Q9JL11	Q9JL11 ratu
19	667.5	9.4	4880	13 Q9JL11	Q9JL11 ratu

20	667.5	9.4	5085	11 Q9UKS6	Q9UKS6 mus musculu
21	667	9.4	6677	5 Q9N435	Q9N435 caenorhabd
22	662	9.4	1229	5 Q94185	Q94185 caenorhabd
23	661	9.3	3507	5 Q23587	Q23587 caenorhabd
24	654.5	9.3	2089	4 Q14676	Q14676 homo sapien
25	653.5	9.3	1514	5 Q9GUW7	Q9GUW7 leishmania
26	651	9.2	1612	5 Q9VY02	Q9VY02 drosophila
27	649	9.2	2768	5 Q9VY02	Q9VY02 drosophila
28	647	9.2	4833	11 Q9VYX6	Q9VYX6 mus musculu
29	647	9.2	5038	11 Q9VYX7	Q9VYX7 mus musculu
30	646.5	9.2	3570	4 Q9VYX2	Q9VYX2 homo sapien
31	635.5	9.0	7962	4 Q10465	Q10465 homo sapien
32	633	8.9	761	10 Q41707	Q41707 vigna ungu
33	632	8.9	632	5 Q17362	Q17362 caenorhabd
34	632	8.9	990	13 Q91803	Q91803 xenopus lae
35	630.5	8.9	971	5 Q9XVS4	Q9XVS4 caenorhabd
36	626.5	8.8	6642	5 Q01761	Q01761 caenorhabd
37	625	8.8	839	2 Q9RX57	Q9RX57 deinoxococcus
38	622.5	8.7	2344	5 Q9N3Y8	Q9N3Y8 caenorhabd
39	615	8.6	801	5 Q23635	Q23635 caenorhabd
40	608	8.6	924	12 Q99307	Q99307 epstein-bar
41	607.5	8.6	379	5 Q27929	Q27929 drosophila
42	605	8.5	1893	5 Q9NKB1	Q9NKB1 drosophila
43	600.5	8.5	409	10 Q9SBM1	Q9SBM1 volvox cart
44	598.5	8.5	956	10 Q9LJ64	Q9LJ64 arabidopsis
45	577	8.2			

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1404 AA.
Q92954	Q92954	Q92954	Q92954	Q92954
AC	Q92954	Q92954	Q92954	Q92954
DT	01-FEB-1997 (TREMBL)	02, Created		
DT	01-FEB-1997 (TREMBL)	02, last sequence update		
DT	01-JUN-2001 (TREMBL)	17, last annotation update		
DE	MEGAKARYOCYTE STIMULATING FACTOR.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]	SEQUENCE FROM N.A.		
RP	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kellerer K., Glanville J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Blood 78:279-279(1991).			
RN	[2]	SEQUENCE FROM N.A.		
RP	Turner K.J., Fitz L.J., Temple P., Jacobs K., Glanville J., Murtha P.,			
RA	Kellerer K., Glanville J., Calveti J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: U0136; AAB09089.1;			
DR	InterPro: IPR000585; Hemopexin.			
DR	InterPro: IPR001212; Somatomedin_B.			
DR	InterPro: IPR002400; GF_cysnot.			
DR	Pfam: PF00045; hemopexin; 2.			
DR	Pfam: PF01033; Somatomedin_B; 2.			
DR	PRINTS: PR00438; GFCYSNOT.			

Db 73 vpsis-ntldtlbpyqlllllslsvnfcrkksqlyk--nfeekfnlasvqsnatqge 129  
 QY 128 KTTSAKENSIEKTSKADLAPTSKVLAKPTKAEETTKGPAITPKPT-----PTPKKP 183  
 Db 130 nsnqnkelinev-keessqlcpv-----lpcqelvtlq--aaapqetvclqepvltleep 179  
 QY 184 ASTPKPEPTPTIKSAPPTPKPAATTTKSAPTTPKEPAATTTKKEPAATTPKEPA-----P 239  
 Db 180 vlt--qepvlt--qepvltqge--vltqgepvltqge--vltqge--vltqgepvltqge 227  
 QY 240 TTKKEPAATTTKSAPPTPKPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA 299  
 Db 228 vltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqge 269  
 QY 300 A-----PTPKKEPAATPKPAATTTKKEPAATTTKKEPA-----PTTKKEP-----SPTTKKEP 347  
 Db 270 vltqgepvltleepvlt--qepvltle--vltqgepvltleepvltqgepvltqgepvltqge 323  
 QY 348 APPTTKSAPTTKKEPAAT-----TTKSAPTTKKEPSPTTKKEPA----- 386  
 Db 324 --vltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqge 380  
 QY 387 -----PTPKKEPAATPKPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA 439  
 Db 381 gnknkksknekkpslesvkqnadegqpldsqisvnaqdsvltqgepvltqgepvltqge 439  
 QY 440 ETAPPTPKKLPTTPEKLAATTTPEKLAATTTPEKLAATTTPEKLAATTTPEKLAATTTPEKLA 499  
 Db 440 lta-----lge--plltqge--vltqgepvltqgepvltqgepvltqgepvltqgepvltqge 470  
 QY 500 NTPKEPA-----PTPKKEPA-----PAPTTKEPAATTTKKEPAATTTKKEPAATTTKKEPA 551  
 Db 471 lltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqge 530  
 QY 552 KPAKELAPTTKKEPAATTTSDKPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA 611  
 Db 531 kswlkfgnknkksknekkpslesvkqnadegqpldsqisvnaqdsvltqgepvltqgepvltqge 590  
 QY 612 KEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA 671  
 Db 591 qep-pltqelaltqge--plltqge--vltqgepvltqgepvltqgepvltqgepvltqgepvltqge 637  
 QY 672 PEPPTTSEVSTPTTKKEPTTIKSPDE-----STPELSAEPITPKA----- 713  
 Db 638 aq-epvltqge--laltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqge 693  
 QY 714 -----LENSPKKEPGVPTTKTTPAATPKPEMTTAKDKTTERDILRTTPE--TTAAPK 761  
 Db 694 lkfgnknkksknekkpslesvkqnadegqpldsqisvnaqdsvltqgepvltqgepvltqgepvltqge 753  
 QY 762 MTKEATTTKKEPTTESKITATTTQVSTTTQDTPPK--ITTLKTTTLAPKVTT-----TK 814  
 Db 754 lltqelaltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqgepvltqge 810  
 QY 815 KITTTTMMKPEETAPKDRATNSKATTPKPK 848  
 Db 811 epstlge--nadekkasegdnlslslseeeteeek 842

Search completed: April 26, 2002, 16:24:09  
 Job time: 439 sec

XX antimycobacterial; antibiotic; vaccine.  
OS Mycobacterium tuberculosis.  
XX WO9741252-A2.  
XX 06-NOV-1997.  
XX 18-APR-1997; 97WO-EP01973.  
XX 29-APR-1996; 96DE-4017184.  
XX (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.  
XX Espitia C, Honisch C, Moreno C, Singh M;  
XX WPI: 1997-549750/50.  
XX N-PSDB; AAT93610.  
XX New DNA and related proteins or RNA derived from M. tuberculosis -  
PT used for diagnosis of mycobacterial infections, monitoring  
PT vaccination and development of anti-mycobacterial agents  
XX Claim 11, Fig 16; 55pp; English.  
PS

This novel 55 kDa protein is encoded by an open reading frame of a *Mycobacterium tuberculosis* DNA fragment (see AAI993610) containing polymorphic GC-rich sequences. Its amino acid sequence shows a high proline content, but there is no homology to any known proline-rich antigens of mycobacteria. Novel *M. tuberculosis* proteins (see AAI31551-57) are claimed. These can be produced as recombinant proteins, especially in bacterial, yeast, fungal or higher eukaryote host cells, and used for diagnosing tuberculosis and other mycobacterial infections in humans or animals. The claimed proteins can also be used for epidemiological studies, for monitoring vaccination, and for the development of vaccines and anti-mycobacterial drugs.

**SQ Sequence 572 AA;**

Query Match	9.38;	Score 612;	DB 18;	Length 572;
Best Local Similarity	30.68;	Pred. No. 5.5e-29;		

```

Qy 177 pPpKEBpASTpKKEPTpTTKSAPpTtKE-----pApTTKSAPpTtKEKpApTTtKRpAp 231
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 pApKsKpAspStrpAp---pmpApImefpIppvpPpDpIsKtCpApPpApPpIppApvp 57
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 232 tPpKpEpApTtTtKEApTTtKSAPpTtKKEpAp-----tPpKkRkApTtTtKEApApTtPKE 283
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 IppvpPpI-----pvpPnkIppAp-----pApPvAvAaVlApCcPpIppIppnhPpApAa 107
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 284 pPpTTpKKEApTTtKKEApTtTtKEApVtApKkRkAp-----tT----- 319
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 pvpGvpIapIppShpAp--pSapvpGvpIapIpsIgrpAvvWkgsfItIstfcrcvsg 165
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 320 -----pKEA-----pTtPKEApApTTtKEBSpTtTtKEApTTtKSAPpTtKEApTt 366
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 evIaGApImpsIrsIspIttCpApI-----pPpIppIppIppntAvPpI--pPpIppv 217
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 367 tKSAPpTtK--EpSpTtTKEP--ApTtPKEApApTtPpKkRkApTtTtKEApApTtTtK 423
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 tAlApPpIppIapIpsIspvPpApPpIpppKkwTtCpPpIapApPpK--tPvVlppgsCpPse 276
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 424 kRApTApKKEApApTtTtKEApTTtPKkLpTtPb--kLApTtPKEApApTtPpELApTtPpP 481
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 kPnpAp--pEpPpKsSpAlppApPpAsmpSavrvPpSpIppAp-----pAapras 327
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 482 tPpTtPpEApApTtPpK--AApTtPKEApApTtTtKEApApTtTtKEApApTtPpK--tEApTtPpK 533
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 mPAlp--pApSpPpApIrcPpIppSpApPp--pApAp---pTppkIlsanpCpPv 379
    | | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY      536  GAAPTLEKPAPTPKKPAKELAPTTTKEPTSTTSOKPAFTTKGAAPTTPKEAPATTP  595
           ||| : ||| ||| ||| ||| ||| ||| ||| ||| : |||
Db      380  ppaap--rppapap--papap|-----ppapdp-----prrpvaanspap  416

OY      596  KEPAFTPKGAPTTLEKPAFTPKKPAKELAPTTKGP-----TSTTSOKPAFTTP  648
           ||| : ||| ||| ||| ||| ||| ||| ||| ||| : |||
Db      417  --ppappappsalp-fvnpappprppaapksrpalpaappappapvratctpppappappap  473

OY      649  KETA--PTTPKEPAP---TTTKKAPPTTPETTPPTTSEVSPTTKKPTTIHNSPDESTP  703
           ||| : ||| ||| ||| ||| ||| ||| ||| ||| : |||
Db      474  nsmaiprpapdpdppllattpapapapppjmsppapapppaapdpapppllindpp---sp  530

OY      704  ELSAEPTEKALENSPEKGEVPTTKPATKP  734
           ||| : ||| ||| ||| ||| ||| ||| ||| : |||
Db      531  pla-----pyvgpaplaplplindrp  549

```

RESULT 15  
AAP60570

AC AAP60570;

DT 24-AUG-1991 (first entry)

Sequence of the Falciparum Interspersed Repeat Antigen (FIRA).

KW Malaria vaccine; antigen; epitope.

OS *Plasmodium falciparum*.

PN W08601802-A

PD 27-MAR-1986

PF 11-SEP-1985; 85WO-0006960

PR 11-SEP-1984; 84AU-0007067

PR 10-SEP-1985; 85AU-0047326

PA (HALL-) HALL INST MED RES.

PI Kemp DJ, Anders R, Coppe

DR WPI; 1986-094065/14

XX  
XX

poly(peptide(s) having antigenicity of RESA or FIRA antigens of

PS Disclosure; Fig 7; 55pp; English.

CC The inventors claim a novel DNA molecule which comprises a

sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA

CC plasmodium falciparum malarial infections.

SQ	Sequence	844 AA
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

2

Query Match	8.4%	Score 553.5	DB 7	Length 844
Best Local Similarity	26.7%	Pred. No. 2.7e-25		
Matches	117	Mismatches 359	Indels 209	Gaps 49

```

Qy      19 IQGVSSVVDKNKN--RTKKKPRPKPVYDEAGSGLDNDFKVTYPRDISITQHNKVS5PK 76
       : : | | : : | : : : : : : : : : : : : : : : : : : : : : : :
Db      14 vdeysltsmenpqtatltnltpdqalsalnahlpr-einhetltsltltnmeneyml 72
Qy      77 TTTAKPILNRSLPNSS-----DYSKETSLTVNKETTYETKETTNNKQHSIDQKE 127

```



```

Db      638 paapppppppvratp-----ppapppppapn 665
QY      839 SKATTPKPKP-----TKAPKPTSTKKPTMPRVAKKPTTPPKRMSTM-----PELNP 889
Db      666 smalpapppppplpaleppppppplpmsppppplpapaadppppplltlmgpspplap 725
QY      890 TSRIEAMHLQTTTRPNQTPNSKLVENPKSEDAAGAGGCT 929
Db      746 vpgaplaplplngirpvrknsll-----gsssgdt 756

RESULT 13
AAV53666
ID      AAV53666 standard; Protein: 4412 AA.
AC      AAV53666;
XX      22-FEB-2000 (first entry)
DT      XX
XX      Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.
DE      XX
XX      Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KW      bone development; g1/1017427/emb/CAA62189.
OS      Unidentified.
XX      XX
XX      W09960164-A1.
XX      25-NOV-1999.
PE      14-MAY-1999; 99WO-US11066.
XX      15-MAY-1998; 98US-0085673.
PR      (OVAR-) QUARK BIOTECH INC.
PA      Elnat P, Mor O, Skalter R, Feinstein E, Faerman A;
PI      WPI: 2000-053304/04.
DR      PT
PT      Identification of stress induced genes for determining risk and
XX      preventing, treating or controlling osteoporosis
XX      PS
XX      Claim 32: Fig 6A-R; 308pp; English.
CC      The present sequence is obtained from a clustral X alignment with
CC      protein 608. Protein 608 was identified using the method of the invention
CC      after subjecting rat osteoblasts to mechanical stress. Expression of the
CC      608 gene was found to be upregulated by about 3-fold in cells subjected
CC      to mechanical strain. The specification describes a method for the
CC      identification of genes responsive to a specific mechanical stress. The
CC      method comprises applying the mechanical stress to an organism (tissue
CC      or cells comprising bone cells), isolating the specific cellular
CC      fractions and extracting mRNA from them, and differentially analysing the
CC      mRNA in comparison with control samples. The method is used to identify
CC      genes whose expression is responsive to a specific stress. The identified
CC      genes are employed in determining risk associated with a physiological or
CC      disease state. The risk determination methods are used for testing a
CC      medication for gene therapy. These medications, or genes identified by
CC      the method of the invention, are used for treating, preventing or
CC      controlling a physiological or disease state (especially osteoporosis or
CC      bone density or other factors causing or contributing to osteoporosis or
CC      its symptoms or other conditions involved in mechanical stress or its
CC      lack. The methods can also be used for advancing research or studies in
CC      bone development.
SO      Sequence 4412 AA;

```

Query Match 9.5%; Score 625.5; DB 21; Length 4412;  
 Best Local Similarity 26.2%; Pred. No. 8.3e-29;  
 Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;

```

QY      123 TDGKEKTSAKENQSIK-----SAKDLPITSVLAKPTPKAETTTKGPALT----- 170
Db      3515 tdfirgllqrvveehvekvhrvleefeeeevfeakpapp-----kpselsekllp 3566
QY      171 TPKEPTPTTP-KEPATTPKEEPTPTTIKSAPTTPEP-APTITKSAPTTPEPATTTTKE 228
Db      3567 pkkpplkvvrkpeppakvpevkkilveekvrvpeepvvpkvpelvpkvevp----- 3621
QY      229 PAPTTPKEPATTTTKEPATTTTKSAPTTPEPATTTPK-----PAPTTPKEPATTTPEK 284
Db      3622 ----ekkvvpvppakkeap-----pkkvpeapkevpkvpvppkvppevpkvpkevp 3672
QY      285 TPPTPEPATTTTKEPATTT-----KEPATPKKAPTTTPEPATTTPEPATTTTKE 337
Db      3673 kaavpekvvpealppkpespppevfedepeesapkp-kpevpvrvvpevpkevpkekv 3731
QY      338 EPSPTTPEPATTTTKSAPTTTKEPATTTTKSAPTTPEEPSPTTPEPATTTPEPATTT 397
Db      3732 paap--pkkpevtvkv-vpeapkevpkekvvpv-ppkkevpptk-----vpevpvav 3782
QY      398 PKKPAPTTPEPATTTTPEPATTTTTPKEPATTTTKEPATTTTKEPATTTTKEPATTT 453
Db      3783 pekvp-----ealppkpespppevfedepeavaleeppaeveeppaapqvtvpknvp 3838
QY      454 PEKLAPTT-----PEKAPTTPEELAPTTPEEPTP-TTPEEPATTTPKAAPTTPKEPAT 507
Db      3839 pekkaavavakkelvpvkvpevpkevpkevpvlpvvpk--peappakvpevpkevpke 3896
QY      508 ---TTPEP-----APTPEKEPATTTTKEPATTTTKEPATTTTKEPATTTPKAPKEL 558
Db      3897 kkvavpkkpevpvppakvpevpkvpv-----leekpavp--vpeaespppevyee--peel 3947
QY      559 APTTKEPTSTSDKAPTTPKGTAPTTPEPATTTPEPATTTPKGTAPTTTKEPATTT 618
Db      3948 ap---eeelapeekxpyvae-----eepevpvpaapeepkklpekvvp-vlkkpeapp 3999
QY      619 PKKPAPTELAPTTTGTPTSTSDKAPTTPKETA-----PTPEKEPATTTPKRP--AP 669
Db      4000 pkepepek-----lekplkprppppppppkedvkekllqkaidkpkvpenpypkeve 4056
QY      670 TPTEPTPTTSEVS--TPPTTKEP-----TTIKSPDESPTPELSAETPALNSKE 720
Db      4057 lrlpikvpggekkvrlklprrkpepkeevlksvlkrrpeeeekve-----pkle-kvkk 4111
QY      721 PGVPTTKPTAAKRP-----EMTTAKDKTTERDLRTPTETTTAAAPKMTKETATTTKETTES 776
Db      4112 pavp---eppppkpveevpvtvtrerkllpeklvpelkpalplpapepdkpkea----- 4164
QY      777 KITATTTQVTSSTTODTTPFKITTTTTLKTTTTLAPKVTTKTTTTEIMNK-----PEETA 830
Db      4165 -----evklkppvpeepclpavvclvpvvgkkaeakpkeaaa 4204
QY      831 KPDRAATNSKATTPKPKOKPTAKPKKPS--TKKKTMPRVKRPKPTTPTP-----RKMTST 883
Db      4205 kpxgplkvypktpspilaerklrlpssggekppdeapflqllkavplkfvtelkdlilt 4264
QY      884 MELNPTSRIAEAMLQTTT 902
Db      4265 esefvgssafefclvpspt 4283

RESULT 14
AAW31855
ID      AAW31855 standard; Protein: 572 AA.
AC      AAW31855;
XX      27-APR-1998 (first entry)
DT      XX
XX      Mycobacterium tuberculosis 55 kDa protein.
DE      Tuberculosis; mycobacteria; infection; diagnosis;
XX      XX

```







PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 23-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0139899.  
 PR 24-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142154.  
 PR 06-JUL-1999; 99US-0142055.  
 PR 08-JUL-1999; 99US-0142390.  
 PR 09-JUL-1999; 99US-0142803.  
 PR 12-JUL-1999; 99US-0142920.  
 PR 13-JUL-1999; 99US-0142977.  
 PR 14-JUL-1999; 99US-0143542.  
 PR 15-JUL-1999; 99US-0143624.  
 PR 16-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 19-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 22-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 27-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 04-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 05-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 23-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151330.  
 PR 01-SEP-1999; 99US-0151330.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155113.  
 PR 23-SEP-1999; 99US-0155138.  
 PR 24-SEP-1999; 99US-0155486.  
 PR 28-SEP-1999; 99US-0156559.  
 PR 29-SEP-1999; 99US-0156458.  
 PR 04-OCT-1999; 99US-0156596.  
 PR 05-OCT-1999; 99US-0157117.  
 PR 06-OCT-1999; 99US-0157753.  
 PR 07-OCT-1999; 99US-0157865.  
 PR 08-OCT-1999; 99US-0158029.  
 PR 12-OCT-1999; 99US-0158232.  
 PR 13-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 21-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 25-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 26-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 28-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 29-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match

14.9%; Score 981; DB 21; Length 763;



```

XX OS Homo sapiens.
XX XX
XX XX Key Location/Qualifiers
XX XX Misc-difference 393..396
XX XX FT /note= "unspecified amino acids"
XX XX FT Misc-difference 444..446
XX XX FT /note= "unspecified amino acids"
XX XX
XX XX WO9523861-A1.
XX XX
XX XX 08-SEP-1995.
XX XX
XX XX 06-MAR-1995; 95MO-CN00015.
XX XX
XX XX 04-MAR-1994; 94CN-0112066.
XX XX
XX XX (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX XX
XX XX Gu X, Han Z, Shen Q.
XX XX
XX XX WPI: 1995-320576/41.
XX XX
XX XX N-PSDB; AAT04546.
XX XX
XX XX New haematopoietic cell growth factor - used for treating
XX XX thrombocytopenia and hematocytopenia
XX XX
XX XX Example; Page 23; 36pp; Chinese.
XX XX
XX XX This sequence represents the human megakaryocytopoietin (MPO) protein.
XX XX This sequence was purified using a carrier which can couple wheat germ
XX XX agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX XX AAR00039 and AAR00040) were used to produce the amplification primers
XX XX shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX XX can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX XX The MPO cDNA can then be inserted into a plasmid which is used to
XX XX transform cells to produce MPO. The MPO sequence is capable of promoting
XX XX colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX XX and stimulating the proliferation of multipotential stem cells. The
XX XX factor may be used for treating thrombocytopenia and hematocytopenia.
XX XX The purification method can be used to isolate MPO from human urine or
XX XX serums of patients with aplastic anaemia, and from animal blood or urine
XX XX by radiation exposing the animals to induce aplastic anaemia.
XX XX
XX XX Sequence 452 AA;

```

```

Query Match 21.3%; Score 1397.5; DB 16; Length 452;
Best Local Similarity 63.2%; Pred. No. 4.8e-76;
Matches 287; Conservative 3; Mismatches 21; Indels 143; Gaps 2;

```

```

QY 1 MAAKTLPIYLLLSVIVIOVSS----- 24
DB 1 MAKKTLPIYLLLSVIVIQVSSQGLSSCAGRGEGYERDAtcncdyncqymccpdcf 60
QY 25 ----- 24
DB 61 kvctaelsockgrcfesfgrccddagckkydcocpdyesfcaevnmpspsskka 120
QY 25 ----- 24
DB 121 ppsgasqtlkstkcrsypknkkkkyseellevdnknrkkrkprkpyvdaag 180
QY 49 SGLDNGDFVYTPDSTTOHNKSVSPKITAKPINPPSPSPNSDSKESLTVNKEET 108
DB 181 agldngdfkvttcstcqhkvssplttackpnpnpslppnsdskesltvnkett 240
QY 109 VETKETTNNKOTSDGKEXKTTSAKETOSIKETSAKDLAPTSVKYAPPTPAETTTGPA 168
DB 241 vetketttnqyscdgkextsketstaketsakdlaptsvklakptpktaetttgpa 300
QY 169 LTTPEKPTPTPKBPASTPTPKPTPTTKSAPTTPKBPAPTTSATTTTKBPAPTPTTKE 228

```

```

DB 301 lttpeptptpkbpasttpkbpptltksaptpckepappttkbpappttk 360
QY 229 PAPPTPKBPAPTPTTKBPAPTPTTKSAPTTPKBPAPTPTTKKBPAPTPTTKBPAPTPTT 288
DB 361 papptpkbpappttkbpappttkshpplprscxxxxctcp-----tpckepbppl 409
QY 289 PREPAPTTPKBPAPTPTTKBPAPTTPAKKBPAPTTPKE 322
DB 410 prshpckepaptpckepaptpakpkaplpplpe 443

```

```

RESULT 8
AAM24516
ID AAM24516 standard; protein; 5179 AA.
XX
XX AAM24516;
AC
XX
XX 12-OCT-2001 (first entry)
DE
XX C899P predicted amino acid sequence.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX immunogenic; gene therapy; vaccine; colonic cancer.
XX
XX Homo sapiens.
XX
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
XX 10-JAN-2000; 2000US-0480321.
XX 13-FEB-2000; 2000US-0504629.
XX 06-MAR-2000; 2000US-0519444.
XX 19-MAY-2000; 2000US-0575251.
XX 29-JUN-2000; 2000US-0609448.
XX 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIYA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -
XX
XX Claim 2; Page 446-462; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them: (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patient's own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples

```

OY	505	PAPTRPEAPATTPKKEPAPTTPKETAPTTAPTKGATTTTLKEPAATTPKKAPAKELAAPTTK	564
Dd	441	papit-kepapit-kepapit-kepapit-kepapit-kepapit-kepap-----tk	488
OY	565	EPISTSDKRAPPTPKATAPTTKEPAPTPKBPAPPTPKGAAPTTLKBPAPPTPKKAP	624
Dd	489	eapatlke-papit-kepapit-kepapit-kepapit-kepapit-kepapit-kepap	541
OY	625	KELAPTTKGESTTSDSKPAPPTPKETAPPTPKBPAPTTPKRAAPTTPPTPTSVEST	684
Dd	542	-----tlkepapitke-papit-kepapit-kepapit-kepapitke-papitkepa-	589
OY	665	PTTYKEPTTIHKSDESPELSAPETPKALENSPKEGVPTTKTPAATKPEMTTAKDKT	744
Dd	590	-pttkepatlkepapit-----kepap-----tlkep-apltkepapitkepapitke--	636
OY	745	TERRDRTTPPETTAAPKKKTETATTTTEKTSKITATTTOVTSNTODTTPFKITLTKT	804
Dd	637	-----papit-----tlkepapit-----kepapitkepapitkepap-----	667
OY	805	TLAPKVTTTKITTETEIMNKEEETAKEBRATNSKATTPRQOKPYAKPKFISTKPKT	864
Dd	668	-----tlkepapitkepapit-----tlkepapitkepapit-apl-tkep--	686
OY	865	MPRVKRPKTPPPRKKTSTMPELANPTSRAPAMLOTTTRPNOTPSKLVEVPKSEDAG	924
Dd	687	-----apttpektmtstmpelnpstrlaeaml--ttctpnqtpnslkvenpkpsedag	739
OY	925	AEGETHMLLRPHVFMEVTPDMXDYLPRVNOGIITINPMLSDETINCGKRVDDGLTLRN	984
Dd	740	aegelpmllrphvfmevtpdmxdylprvpngiilnplmsdetlnicgkpydgiltlrn	799
OY	985	GTVLAERGHAFWMLSPPSPSPARRITTEWGGISPIDVTFRNCNEGKTFEFKDSQYWR	1044
Dd	800	gtlvafgrghfwmllspispsparilltevwglpsldvtfrtcneogkffffksqywr	859
OY	1045	TNDIKDAGYEPKPIFGGGLTGQIVAAALSTAKYKNPESVYFFK	1088
Dd	860	tndikdagypkpifgfgltgqivaalsta-yknpesvyffk	902

RESULT 6  
 AAB60569  
 ID AAB60569 standard; Protein: 472 AA.  
 XX AC  
 XX AAB60569;  
 Df 27-APR-2001 (first entry)  
 DE Bovine MSF orthologue, superficial zone protein (SZP).  
 KW Bovine; CACP protein; campodacyly-arthropathy-coxa vara-pericarditis;  
 KM Superficial zone protein; SZP; MSF orthologue; synovial lubricant;  
 OS osteoarthritis; joint lubrication; osteopahlic; antiarthritic.  
 PA Bos taurus.  
 PN WO200107068-A1.  
 PD 01-FEB-2001.  
 PF 21-JUL-2000; 2000WO-US20002.  
 PR 23-JUL-1999; . 99US-0145328-  
 DR 19-JUL-2000; 2000US-0145328.  
 (UYCA-) UNIV CASE WESTERN RESERVE.  
 WPI; 2001-182721/18.

New composition comprising the campodactylly-arthropathy-coxa  
 para-pericarditis protein in combination with an anesthetic, useful for  
 treating osteoarthritis, or as lubricants of tissue and joints -  
 Example 1; Fig 4; 34pp; English.

Query Match	26.0%;	Score 1707.5;	DB 22;	Length 472;
Best Local Similarity	79.9%;	Pred. No. 1.5e-94;		
Matches 319;	Conservative 24;	Mismatches 53;	Indels 3;	Gaps 3.

Sequence	472 AA;
SQ	

New composition comprising the campodactylly-arthropathy-coxa vara-pericarditis protein in combination with an anesthetic, useful for treating osteoarthritis, or as lubricants of tissue and joints -

Example 1; Fig 4; 34pp; English.

The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the campodactylly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder campodactylly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents the bovine orthologue of human MSF, superficial zone protein (SZP).

OY		832	PKDATNSKATTPPKQPTKA PKKPTSTKKPKTWPVRVKRTPTPRK -MTSNMPLNPT	890
Dd		76	pkratnsqvttppkpkpkatkpkrt-privrkptcptppkttsamptpc	134
OY		891	SRIAPAMQTTRRNOPNPSKLIVPNKSESDAGAEETPHMLLRPHVMEPVTPDMXYL	950
Dd		135	s-lpeamqctttrprnpnseidvnsenedgdaaegekphmlfrppvltprlvpgetll	193
OY		951	PRVPNGGIIINPMSDETNIENCKRPVDGLTLRNGSLVAFRGHFEMWLSPSPSPSARRI	1010
Dd		194	vrgpqsgfgjnmfedeelnmgrrvdvdltlrlngilvafrghfemwltpctpppprrl	253
OY		1011	TEWNGIPSPIDIVFTRCNCEGTFFFKSOXWRFTNDIKDAGXPKRPIFGGILTGQIVA	1070
Dd		254	tewngipspidivftrcncegtfffksgywrftndikdagypkliskgfglnqkiva	313
OY		1071	ALSTAKYKNMPESVYFEFKGGSIQOYIKQEPYOKCPGRPALNPVYGEMTOVARRRRE	1130
Dd		314	alslakysrpesvyffkrgsvvgqlkyqdeplqctgrpalnpsvgetaqvrrrfc	373
OY		1131	RATGSGOPTIRIOYSPARLAYODKCVLHNEKVSLMRGLDPNVYTSALSIPNIEKPGY	1190
Dd		374	rategqvhtirihyprvpygdqkfllmekvslwlgldpnvvtsalstpnitkpgy	433
OY		1191	DYYAFSKDYNNIDVPSRARATTSSGOTLSKWNCPC	1229
Dd		434	dyyafskdyynnidvpsrarattssgotlskwncpc	472
RESULT		7		
AAR80041		ID	AAR80041 standard; Protein; 452 AA.	
XX		AA	AAR80041;	
XX		DT	10-Apr-1996 (first entry)	
XX		DE	Human megakaryocytopenulin protein.	
XX		KW	Human; megakaryocytopenulin; wheat germ agglutinin; heparin;	
KW		KW	megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;	
KW		KW	multipotential stem cell.	





```

Db      |||
721 apttpkpkapkelapctkeptsdsdpapctipkagtapctpkcapctpkcapctpkp 780
QY      |||
606 TAPPTTKEPAPTTPPKAPKAPKELAPTTTGTGPTSTSDKAPAPTPPKETAPTPKEAPPTPK 665
Db      |||
781 tapctpkcapctpkpkapkelapctkxptsdsdpapctipkcapctpkcapctpkp 840
QY      |||
666 KPAAPTPPEPPTTSVSPPTTTPKPTTHKSPDESTEPKELSAEPPTKALENSEKPEGVP 725
Db      |||
841 kpaaptppeppttsvsppttkeptclhkspdesstepelasepckalensekpegyvt 900
QY      |||
726 TKTPAATKPEMTTAKDKTTERDLRTPTPTTAAPKMKETATTTTEKTESTTATTOV 785
Db      |||
901 tkpapatkpeamtakdktterdlrtptpettaapkmketaetketskitacltq 960
QY      |||
786 TSTTTQDTTFFKTTTTLKTTTLAPKVTTKTITTTTEIMKPEETAKPKDRATNSKATTPK 845
Db      |||
961 tssttqdtffktttllktttlapkvttkctltteimkpeetapkdratnskattpk 1020
QY      |||
846 POKPTKAPKPTSTKPKPTMPRVKPKPTTPPKMTSTMBELNPTSRIRIEMLOTTTRPN 905
Db      |||
1021 pdkptkpkpstkpkpcktmprvkpktpctpkmtstmbelnpstriaeamloqtrpn 1080
QY      |||
906 QTPNSKIVEVNPKSEDAAGEPPEHMLLRPHYMEPEVTPDMDYLPBVPNOGIIINPMLS 965
Db      |||
1081 qtpnsklivevnpksedagaegecpmhlirphvfmpevtpdmqylprvpngqilinpms 1140
QY      |||
966 DETNINCNGKPYDGLTTLNGLTVAFRGHYEMLSFSPSPAPARITPVGISPIDTVP 1025
Db      |||
1141 detnincngkpydglttlrlngclvafrghyfmblspfspaparriltevwipspidvft 1200
QY      |||
1026 RCNCEGKTFEFPKSDQYMETNDIDAGYKPIFKFGGLTGOIYVALSTAKYKNWPESVY 1085
Db      |||
1201 rcncegkttfifkdsqywtfnidkagypkplfifggltsqivaalstakynwpeasy 1260
QY      |||
1086 FFKRGSGIOQYTYKQEPYQKCGRRPALNVPYEGMTOVRRRRFPRAIGPQGTTRIOY 1145
Db      |||
1261 ffrgsgisqylykqepvqkcpgrrpalnvpvygemtvttrrrferralpqpqutitfgy 1320
QY      |||
1146 SPARALAYQDKCYLHNEVKYSILMRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYNTDV 1205
Db      |||
1321 sparlayqdkyvlhnevksilwrglpnvvtasaislpnirpdkdydyafskdyntdv 1380
QY      |||
1206 PSRTARAITTRSGOTLSKYWYNCP 1229
Db      |||
1381 psrtaraittrsgqltskwyncp 1404

```

## RESULT

4  
ID AAM24322 standard; Protein: 1299 AA.

XX AAM24322;

XX 12-OCT-2001 (first entry)

DE Human EST encoded protein SEQ ID NO: 1847.

XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KW diagnostics; forensic test; gene mapping; genetic disorder;

KW biodiversity; gene therapy; nutrition.

OS Homo sapiens.

PN WO200154477-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-0502687.

XX 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

```

PR      03-AUG-2000; 2000US-0611451.
PR      15-SEP-2000; 2000US-0663870.
XX      (HYSE-) HYSEQ INC.
XX      Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI      Cao Y, Drmanac RA, Zhang J, Werhman T;
XX      WPI: 2001-476164/51.
DR      N-PSDB: AAH98981.
PT      Isolated polypeptide for treatment of diseases, diagnostics, raising
PT      antibodies and research use -
PS      Claim 20: page 1198-1201; 1275pp; English.
XX      The present invention provides the protein and coding sequences of novel
CC      proteins from a variety of organisms, including human, dog, cat, horse,
CC      cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC      urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC      from the organism of interest. They can be used in diagnostics,
CC      forensics, gene mapping, identification of mutations, to assess
CC      biodiversity and for nutritional purposes. The present sequence is a
CC      protein of the invention.
SQ      Sequence 1299 AA:
Query Match 89.7%; Score 5894.5; DB 22; Length 1299;
Best Local Similarity 86.5%; Pred. No. 0;
Matches 1120; Conservative 0; Mismatches 0; Indels 175; Gaps 1;
QY      1 MAMKTLPTLYLLLSLVYIQVSS----- 24
Db      1 mawktlplyllllslvfyqysqdlsscagrcgyrdatcndynoghyameccpdf 60
QY      25 ----- 24
Db      61 kvctaelsockrfesfergrecdedagckkydkccpdyesfcaevhnpstpskkap 120
QY      25 ----- 24
Db      121 ppsgaagtlkstktpkpnkkkktkkyseelteesvsengessssssssstlw 180
QY      25 -----VKDKKKNRTRKKRPKPPVVDGSGLDNGDKVTTPTDST 65
Db      181 kksksksaanaelqkklkvkdknkrtkkkpkpvpvdeagsgldngdfvltptdst 240
QY      66 TOHNKYSTSPKITTAKPINDRPSLPNSDTSKETSITVKKETTVEKTTTNKQSTNDG 125
Db      241 tqhkvstspkittakpindrpslppnsdstsketsitlvkettvekttnkqstsdg 300
QY      126 KEKTTSAKEQOSTKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPPTTPKEPAS 185
Db      301 kektsakeqstsktsakdlaptskvlakptpkkaetttkgpalttpkepttpkepas 360
QY      186 TTPKEPPTTITKSAPTTTKEPAPTTPKSAPTTTPKEPAPTTPKAPPTTPKAP 245
Db      361 ttpkepptitksaptttkepapttpkspaptttkepapttpkapaaptttkp 420
QY      246 APTTTSKAPTTTPKEPAPTTPKPPAPTTTPKEPAPTTPKPPAPTTTPKEPAPTTPK 305
Db      421 aptttskaptttkepapttpkppaptttkepapttpkapaaptttkp 480
QY      306 EPAPTAPKKPAPTTPKEPAPTTPKAPPTTTPKEPAPTTPKAPPTTTPKAP 365
Db      481 epaptapkkpapttpkepapttpkapaaptttkepapttpkapaaptttkp 540
QY      366 TTKSAPTTTPKEPAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAP 425
Db      541 ttksaptttkepapttpkapaaptttkepapttpkapaaptttkp 600
QY      426 APTAPKEPAPTTPKEPAPTTPKKLTPPTPEKLAPPTPEKAPAPTTPPEELAPPTPEEPPTT 485

```

[illegible]

QY 1086 FFKKGGSTQOYITTKQEPVOKCGRRPALNYPVGEEMTOVRRRRPERAIGPSQHTTIRIQY 1145  
 |||||||  
 Db 1261 ftkrgsgsqylykqepqkcpgrtpalnpvygemctqrrrrrrferiagpsqhtlrltlyq 1320  
 |||||||  
 QY 1146 SPARLAYODKGVILHNEVSVSLMRGLPMNVVTSALSLPNIRKPDGYDYAASKOYVINDV 1205  
 |||||||  
 Db 1331 sparlaydkgvllhnevsvslmrwglpmnvsvtsalslpnrkpdgydyafakdgyndv 1380  
 |||||||  
 QY 1206 PSRTARAITTSGOTLSKWTNCP 1229  
 |||||||  
 Db 1381 psrtaraittrsgtllskwtynpc 1404  
 |||||||  
 RESULT 2  
 AAB60568  
 ID AAB60568 standard; Protein: 1404 AA.  
 XX  
 AC AAB60568;  
 XX  
 DT 27-APR-2001 (first entry)  
 XX  
 DE Human megakaryocyte stimulating factor (MSF, CACP).  
 XX  
 KW Human: CACP protein; camplodactilyl-arthropathy-coxa vara-pericarditis;  
 KW MSF: megakaryocyte stimulating factor; synovial lubricant;  
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KW anclathritic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200107068-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 21-JUL-2000; 2000MO-US20002.  
 XX  
 PR 23-JUL-1999; 9905-0145328.  
 PR 19-JUL-2000; 2000US-0145328.  
 XX  
 PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX  
 PI Warman MI;  
 XX  
 PT WPI; 2001-182721/18.  
 XX  
 PT New composition comprising the camplodactilyl-arthropathy-coxa  
 PT vara-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints  
 XX  
 PS Example 1; Page -: 34pp; English.  
 XX  
 CC The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the camplodactilyl-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder camplodactilyl-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a Genbank accession number was given. This  
 CC sequence was therefore obtained from Genbank (U70316).  
 XX  
 SQ Sequence 1404 AA;

Query Match 98.5%; Score 6470.5; DB 22; Length 1404;  
 Best Local Similarity 87.5%; Pred No. 0;  
 Matches 1229; Conservative 0; Mismatches 0; Indels 175; Gaps 1;  
 1 MARKTPIYLLLLLSVFVLIQOVS-----24  
 |||||||  
 Db 1 marktpilyllllsvflvqvsqdlsscagrcgysrdactncdynqhyemccpdf 60  
 |||||||  
 QY 25 -----24  
 Db 61 kvotaelockgrcfesfergrecdcaqckkydkccpdyefcaevhnpbpskcap 120  
 |||||||  
 QY 25 -----24  
 Db 121 ppsgaagtlstkrspkpnkkkvlkieseeeltehsvengeessssssstliw 180  
 |||||||  
 QY 25 -----VKDNKKNTKKKPPKPPVVDAASSGIDNDFKTPEDTST 65  
 |||||||  
 Db 181 kikesksaanrelqkklkvdknkrkklkppkpvvdaagsqldngfkvlpdstc 240  
 |||||||  
 QY 66 TQHNKVTSEKITTAKEINRPSLPKSDTSKETSILVNKEETVETKETTNNKOTSDG 125  
 |||||||  
 Db 241 tqhnkvstspkiltakpimprpslppnsdskeslsvnkettvettkqtstgd 300  
 |||||||  
 QY 126 KEKTTSAKETOSIKKTSADLAPTSKVLAAPPKAETTTGPAITTPKEPTTPKEPAS 185  
 |||||||  
 Db 301 kektsakeitgsiektsakdlaptskvlakpkaettkgpaalttkpttkpepas 360  
 |||||||  
 QY 186 TTPKEPTTPKSAPTTPKEPAPTTTSAPTPKEPAPTTTPKEPAPTTTPKEP 245  
 |||||||  
 Db 361 ttpkepttpktsaptpkpeaptpktsaptpkpeaptpkpeaptpkpeaptpk 420  
 |||||||  
 QY 246 APTTKSAPTPKPEPAPTTPKPAPTTPKEPAPTTPKEPTTPKEPAPTTTPKEP 305  
 |||||||  
 Db 421 apttksaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 480  
 |||||||  
 QY 306 EPAPTAPKPAPTTPKPEPAPTTPKPEAPTTTKESPTTPKPEAPTTTPKPEAPT 365  
 |||||||  
 Db 481 epaptapkkpapttpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 540  
 |||||||  
 QY 366 TTKSAPTPKPEPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKP 425  
 |||||||  
 Db 541 ttksaptpkpepttpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 600  
 |||||||  
 QY 426 APTAPKPEAPTTTPKETAPTTPKKLTPTTPKLAPTTPKPEAPTTPEELAPTTPPEPTT 485  
 |||||||  
 Db 601 aptapkeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 660  
 |||||||  
 QY 486 PEEPAPTTKKAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTPK 545  
 |||||||  
 Db 661 peepapttkkaapntpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 720  
 |||||||  
 QY 546 APTTPKKPAPELAPTTTPKEPTSTSDKPAPTTPKGTAPTTTPKPEAPTTTPK 605  
 |||||||  
 Db 721 apttpkkpapelaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 780  
 |||||||  
 QY 606 TAPTLKEPAPTTPKPAPELAPTTTPKGTSTTSKPAPTTPKPEAPTTTPKPEAPTTPK 665  
 |||||||  
 Db 781 taptlkeaptpkpaapelaptpkpeaptpkpeaptpkpeaptpkpeaptpk 840  
 |||||||  
 QY 666 KPAPTPEPTTPPTSEVSTPTTPKPEPTTHKSPDESPELSAPPKALENSKPEGVPT 725  
 |||||||  
 Db 841 kpaptpepttpptsevspttpkpeptthkspdespeelsaepkalenskpegvpt 900  
 |||||||  
 QY 726 TKTPAATKPEMTTAAKDKTTERDLRTTPETTTAAPKMTKETATTPKTESKTTATTTGV 785  
 |||||||  
 Db 901 tktpaatkpemtttaaakdktterdlrttpeetttaapmketatectesktatattgy 960  
 |||||||  
 QY 786 TSTTGOPTPEKTTTLTKTTLAPKVTYTKTTTTEIMNKPEETAPKRDATNSKATPY 845  
 |||||||  
 Db 961 tsttgdtpkltlcttlapkvtttkktitttelmoipeetapkdatskatp 1020  
 |||||||

Qy	25	-----VKDNKNRKTKKKPTPKPPVVDGSSGJDNDEKVTTPDST	6
Db	101	kikseksaanerqlkklvkdnknrtkkkprbpvvdgsgjdnadkvtltpdst	2
Qy	66	TOHNKVSPTITTKAPINRPSLPNSDSKTTSLTVNKEVTVEETKEVTYTNKQTSYDG	1
Db	241	tqhnkvstspkltakpimprpslpnsdstskstsltvnketvcteketttnkstsdg	3
Qy	126	KEKTTSAKETQSIEKTSADLAPTSVLAKPKPAETTKGALVTPKEPPTPKPAS	1
Db	301	kekttsaketqsilektsadlaptsvlakprkaettkgalvtpkepptpkpas	3
Qy	186	TTPEKEPPTTIKSAPTPKEPAPTTTKSAPTPKKEPAPTTTKEPAPTPPKKEPAPTTTKEP	2
Db	361	ttpekeppttiksaptpkepaptttksaptpkkepaptttkepaptpkkepaptttkep	4
Qy	246	APTTTSAPTPPKKEPAPTPPKKRAPTPPKKEPAPTPKEPPTPKKEPAPTPPKKEPAPTPPK	3
Db	421	aptttsaptpkkepaptpkkrapttkkrapcttkkepaptpkkepaptpkkepaptpk	4
Qy	306	EPAPTPKRAPTPPKKEPAPTPPKKEPAPTTTKKEPSPTPKKRAPTTTKSAPTTKKEPAP	3
Db	481	epaptpkkrapttkkepaptpkkepaptpkkepaptpkkepaptpkkepaptpk	4
Qy	366	TTKSAPTPKKEPSPTTKKEPAPTPPKKEPAPTPPKKRAPTPPKKEPAPTPKEPAPTTTKPK	4
Db	541	ttksaptpkkepaptpkkepaptpkkepaptpkkrapttkkrapcttkkepaptpk	60
Qy	426	APTPAKEPAPTPPKETAPTPKLTPPTPKELAPTPPEKAPTPPEELAPTPPEEPPTT	48
Db	601	aptpakkepapttketapttkltpcttkelaptpckrapctpeelapttpeepptt	66
Qy	486	PEEPAPTPPKAAANTPKKEPAPTPPKKRAPTPPKKEPAPTPPKETAPTPKGAPTTLEP	54
Db	661	peepaptpkkaapntpkkepaptpkkrapttkkrapcttkkepapttkgapttlkep	72
Qy	546	APTPPKKAPKELAPTTTKBPTSTSDSKRAPTPPKGAPTPPKKEPAPTPPKRAPTTTKG	60
Db	721	aptppkkrapkelaptttkbptstsdskraptppkgaptppkkepaptpkraptttkg	78
Qy	606	TAPTTLEKAPATTPPKRAPKELAPTTTKBPTSTYSKRAPTPPKETAPTPKKEPAPTPPK	66
Db	781	tapttlkepapttkrapkelaptttkbptstyskraptppketaptpkkepaptpk	84
Qy	666	KRAPTPPEPPTTSEVSTPTTKKEPTTIHNSPDSTBELSAEPPKALENSPKEPVP	72
Db	841	kraptppeppttsevspttkkepttlhnspsdtbelsaepkalenspekvpt	90
Qy	726	TKTPAATKPEMTTAAKDTERDLTPPEPTTAAPKMKIEXATTTTEKTTESKTATATTQV	78
Db	901	tktpaatkpemttaakdterdltppepttAAPKMKIEXATTTTEKTTESKTATATTQV	78
Qy	786	TSTPTODTTPKTIITLKTTLTAAPKVTYTKITTTTELINMKPEELAKPKDAPATNSKATTPK	84
Db	961	tsttcddtpttkitlkttlctlapkvtctkilttelinmkpeelakpkdatsksttpr	102
Qy	846	POKPTAPKRPKPTSTKKPTMPRVKPKTPPKKSTMBELNPTSLIAFAMLTQTTTRN	90
Db	1021	pkpctkapkpkstckpkptmprvtkpktptrpkkstmbeLNPtSLIAFAMLTQTTTRN	90
Qy	906	QPPNSKIVEVMPKSDGAGAGETPPHMLBPHVMEVTPDMQVTLRRVNOGIINPMLS	96
Db	1081	qppnskivevmpksgdagagetrpPhmlbphvmevtpdmqvTLRRVNOGIINPMLS	96
Qy	966	DETINCSKPYDGLTJLRNGILVAFRHHYPMMLSPSPSPAPARTIEVWGISPIDVTFT	102
Db	1141	detinongkpydgtlJtLrngilvafRHHYPMMLSPSPSPAPARTIEVWGISPIDVTFT	1200
Qy	1026	RCNCEGKTFEPKSDOSYARFNNDIKDAGYKPKPIFKGFGJLGOIVALSSTAKYKMMPEVY	108
Db	1201	rcncegkttffkdsqywftrfndikdagypkPlfkGfgJLtgJlvaalsstakvkmmpevy	1260

Mon Apr 29 08:35:13 2002

aal.rag

Page 1.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:50 ; Search time 138.84 Seconds  
(without alignments)  
655.691 Million cell updates/sec

Title: AAI  
Perfect score: 6568  
Sequence: 1 MAWKTLPIYLLILLVSVFVIO.....AAITTRSCQTSKRWYNCP 1229

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /Geneseq\_1101: \*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT: \*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT: \*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT: \*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT: \*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT: \*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT: \*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT: \*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT: \*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT: \*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT: \*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT: \*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT: \*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT: \*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT: \*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT: \*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT: \*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT: \*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT: \*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT: \*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT: \*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6470.5	98.5	1404	13 AAR26049	MSF precursor. Sy
2	6470.5	98.5	1404	22 AAB60568	Human megakaryocyt
3	6470.5	98.5	1404	22 AAB29773	Human megakaryocyt
4	5894.5	89.7	1299	22 AAM24322	Human EST encoded
5	3484	53.0	902	22 AAB29778	Human MSF-derived
6	1707.5	26.0	472	22 AAB60569	Bovine MSF ortholo
7	1397.5	21.3	452	16 AAR80041	Human megakaryocyt
8	1168	17.8	5175	21 AAM24516	C899P predicted am
9	981	14.9	1664	21 AAG38942	Arabidopsis thalia
10	950	14.5	1664	19 AAW3106	C. thermocellum OI
11	772.5	11.8	1325	22 AAW3645	Peptide #2327 enco

12	715	10.9	763	18 AAW31852	Myobacterium tube
13	625.5	9.5	4412	21 AAY53866	Sequence g1/101742
14	612	9.3	572	18 AAW31855	Myobacterium tube
15	553.5	8.4	844	7 AAP60570	Sequence of the Fa
16	542	8.3	807	21 AAY54467	Amino acid sequenc
17	520.5	7.9	788	21 AAY54466	Amino acid sequenc
18	508	7.7	1837	21 AAB11726	Cryptosporidium pa
19	506.5	7.7	744	9 AAP82975	Bioadhesive precu
20	496.5	7.6	2971	21 AAB41231	Human OREX ORF95
21	489.5	7.5	2972	22 AAB50362	Human SRCAP. Homo
22	489.5	7.5	3118	22 AAB50362	Portion of Cryptos
23	489	7.4	1721	21 AAB11727	P. yoelii SSP2 ant
24	488.5	7.4	826	13 AAR26042	Peptide #2892 enco
25	488	7.4	617	22 AAM16458	Peptide #2869 enco
26	488	7.4	617	22 AAM04187	Human MUC11 polype
27	488	7.4	957	21 AAY59288	C900P predicted am
28	488	7.4	957	22 AAM24513	Cryptosporidium pa
29	488	7.4	1127	19 AAM48299	Human protein sequ
30	485	7.3	1127	22 AAB95541	Bioadhesive precu
31	476.5	7.3	652	9 AAP82974	Peptide #1317 enco
32	467.5	7.1	511	22 AAM14883	Peptide #1289 enco
33	467.5	7.1	511	22 AAM27312	Peptide #1289 enco
34	467.5	7.1	511	22 AAM02607	Peptide #1289 enco
35	450.5	6.9	378	12 AAR14160	Peptide #1289 enco
36	446.5	6.8	378	12 AAR14162	Peptide #1289 enco
37	446.5	6.8	378	12 AAR14160	Peptide #1289 enco
38	443	6.7	2870	21 AAY95559	Caenorhabditis ele
39	443	6.7	3178	21 AAY95559	Caenorhabditis ele
40	442.5	6.7	3178	21 AAY95559	Caenorhabditis ele
41	435.5	6.6	2819	22 AAB35408	Human O76G27 gene
42	431.5	6.6	2665	22 AAM14533	Peptide #967 enco
43	431.5	6.6	2665	22 AAM26950	Peptide #987 enco
44	431.5	6.6	2665	22 AAM02259	Peptide #941 enco
45	431.5	6.6	3266	21 AAB42491	Human ORFX ORF2255

# ALIGNMENTS

RESULT 1	AAAR26049	standard; Protein; 1404 AA.
XX	AAAR26049:	
XX	02-FEB-1993 (first entry)	
XX	MSF precursor.	
XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;	
KW	stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	Synthetic.	
XX	Location/Qualifiers	
FT	1..26	/label= Exon_I
FT	26..67	/label= Exon_II
FT	67..107	/label= Exon_III
FT	107..157	/label= Exon_IV
FT	157..200	/label= Exon_V
FT	200..1141	/label= Exon_VI
FT	1411..1166	/label= Exon_VII
FT	1166..1212	/label= Exon_VIII
FT	1213..1266	/label= Exon_IX
FT		

CC factor, ie. one or more of these exons may contain sequences which  
 CC direct proteolytic cleavage, adhesion, organisation of the cellular  
 CC matrix or extracellular matrix processing. Both naturally occurring  
 CC and non-naturally occurring MSF's may be characterised by various  
 CC combinations of alternatively spliced exons from this sequence, with  
 CC the exons spliced together in differing orders to form different  
 CC members of the MSF family.

XX Sequence 111 AA;

Query Match 8.4%; Score 613; DB 13; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-28;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 LSSCAGRCGEGYRDATCNCQYHMECCPDFFKRYCTAELSCKGRCESPFERGREGDC 86  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 8 lsscgagrcggyrdatcncqyhmeccpdkrvctaelsckgrecsferygrecdc 67  
 QY 87 DAQCKKYPKCCPDYEFCAEYHNPSTSPSSKKAAPPSSGASQTIK 130  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 68 daqckkydkcpdyefcaevhnpstspsskkaappsgasqtlk 111

Search completed: April 26, 2002, 16:25:34  
 Job time: 524 sec

Query Match	8.88;	Score 643;	DB 21;	Length 4412;
Best Local Similarity	24.68;	Pred. No. 9.9e-26;		
Matches 272;	Conservative 115;	Mismatches 485;	Indels 234;	Gaps 51;
OY	ELSCGRCFES-----FRCGRCDCAOCCYKDCCPYESC-----EVHNPTSPSS	116		
Db	3278 qlschvgseplrlqwlkagre-----lkpdcrcfsfaagtlavlrlrdvakadsgdy	3331		
OY	117 KKAPPGSASOTIKSTTRSPKPPKKTKKYIEEELTEVNDKNNRKKKPT-----	170		
Db	3332 cksasvag-sdltkskvlikdkpavapakkaavdgrrlffvsepslrvektlatflak	3390		
OY	171 ----PKPVNDAGS--GIDNG-----DFKVTTEEDTST-----QHNNK	203		
Db	3391 vggdipmvtkqkvrqlngdgrryflingkdgaeklelrdtktsqglyrcvafnehgsl	3450		
OY	204 STSPKITTAK-----PINRPSLPPNSDTSEISLVYNNKETVETKETTTTN	250		
Db	3451 esnvnllgyderkkgeklieglrdamlkkrlpilkkgaseeeldl-mellknvdpkeykya	3509		
OY	251 KOTS--TDCKEKRTSKSEKQSIKST-----SAKDIAPISTKVLAKPTPKAETTKKCPALT	302		
Db	3510 rmygltldrlglrrveehrvkvhrlvfeaaeevrekapp-----kgsels	3561		
OY	303 ----TPKEPPPTTP--KEPASTPKPEPPPTTIKSAPTTKP--APTITTSAPTTKEPAP	355		
Db	3562 ekllppkprtkvnpkrpeppakvpeppkllvveekrvpeeprrvpcltkvpevlprkevvp	3621		
OY	356 TTTKEPAPTTKEPAPTTTKEPAPTTTSAPTTTKEPAPTTTK-----PAPTTKKEPAPT	411		
Db	3622 -----ekvvpvppakkpeap-----pkyvpeepkevpekkkvyppppkkvpeppk	3667		
OY	412 TPKEPPTTPKPEAPTTTKEPAPTTP-----KEPAPTAAPKKRAPTTKEPAPTTPKEPA	464		
Db	3668 vpevpkaavpekkvpealprkpespppevfespeesaprk--kpevprrvpevpkev	3726		
OY	465 PTTTKEPSTTTKEPAPTTTSAPTTTTRPPAPTTTSAPTTKREPSPTTTKEPAPTTPK	524		
Db	3727 pekkvpaap--pkkvevrvk--vpeapkevvppekkvvp--ppkkvpeppk-----vpe	3777		
OY	525 PAPTPPKRAPPTPKPEAPTTTKEPAPTTTTRKPPAPAPKEPAPTTTKEPAPTTPKKLT--	582		
Db	3778 pkvavpekkv-----eaiprkpespppevfespeesaloeppveevepepaarpqvtlv	3833		
OY	583 --PTTPEKIAPTT---PEKPAPTTPEELAPTTPEEPTP--TTPEEAPTTPKKAANTRK	635		
Db	3834 pknpvpekkapavakkpelpvkvpevpkevpekkvvlvvpk--peaprkavpevpk	3891		
OY	636 EAP-----TTPKEP-----APTTPKEPAPTTTKEPAPTTPKSTAPTTTKEPAPTTPKP	685		
Db	3892 evvppekkavnpkpevpapkvpevpkkv---leekpavp---vpepeesppveyee	3943		
OY	686 ABRELAAPTTKEPTSTSDKPAPTTTGGAPTTPKPEAPTTTKEPAPTTTKEPAPTTTGAAPTTLKE	745		
Db	3944 -peelap---eeelapeekvypvae-----eeevppvapeppklllpekkvvp--vik	3994		
OY	746 PAPTTTTRKRAPKELAPTTTGGTSTSDKPAPTTTKEPETA-----PTTPKEPAPTTTPK	798		
Db	3995 peappkpepekev---leekpklrpppppapepvedvkekllfqalrkakvpenpvy	4051		
OY	799 P--APTPEPTPEPTTSEVS--PTTTTKEP-----TTTHKSDESTPELSAETPKALE	847		
Db	4052 pekvellrlkvpbgsekkvklrlperpepkeevvlksvlrlkrpeeeekve---pkkle	4107		
OY	848 NSBKEGCVPTTKTAPATKP-----EAMTTAKDKTTERDLRTTPETTYAPAKMTKETATTE	903		
Db	4108 -kvkpvavp---epppkpveevevplvltkrcklpeklvpeklpailpdpapepdkpe	4163		
OY	904 KTTESKITATTTODTTPFKITTLTKTTTLAKVTTTKTITITTEIMNK-----	957		
Db	4164 a-----evktlkpppvpeepcliaaprvlvpyvgkkaeakap	4199		
OY	958 PEETAPKORATNSKATTPKPOKPTYAKPKPTS--TKKPTMBPVRKKKPTTPP-----R	1010		

Dd	4200	keeaakpkipikvnpktpspieaarrrkrlppsgagskppdeapfeylkaavlplfvteik	4259
Oy	1011	KMTSTMPELNPFSKIAEMQTTT	1034
Dd	4260	dilltesefvgssallfeclvpst	4283
RESULT 15			
ID	AAK26050	standard; Protein; 111 AA.	
XX	AAK26050;		
AC			
DT	02-FEB-1993	(first entry)	
XX			
DE	MSF-K130.		
XX			
XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;		
RW	stability; proteolytic cleavage; adhesion; alternative splicing,		
KM	MSF-precursor.		
KM			
XX	Synthetic.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	Region	1..4	
FT		/label= Enterokinase_site	
FT	Region	7..48	
FT		/label= Exon_II	
FT	Region	48..88	
FT		/label= Exon_III	
FT	Region	88..111	
FT		/label= Exon_IV	
PN	WO9213075-A.		
XX			
PX	06-AUG-1992.		
XX			
XX	17-JAN-1992;	92WO-US00433.	
PF			
PR	18-JAN-1991;	91US-0643502.	
PR	10-SEP-1991;	91US-0757022.	
XX			
PA	(GENY ) GENETICS INST INC.		
XX			
PI	Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;		
XX			
DR	WPI; 1992-284660/34.		
DR	N-PSDB; AAO27224.		
XX			
PT	New human mega-karyocyte stimulating factors - for treating		
PT	immune deficiencies; cancer; exposure to radiation or drugs,		
XX	bacterial and viral infections, etc.		
XX			
PS	Disclosure; Fig 3; 87pp; English.		
CC			
CC	The sequence covered by exons II, III and IV encodes megakaryocyte		
CC	stimulating factor (MSF). This sequence was used to produce MSF as		
CC	a fusion protein with thioredoxin. This sequence was derived from a		
CC	MSF-precursor (see also AAK26050).		
CC	The precursor cDNA sequence contains sequences derived from human		
CC	megakaryocyte colony stimulating factor (meg-CSF). Exon I contains		
CC	the initiating methionine, and encodes a classical mammalian protein		
CC	secretion signal sequence. The sequence encoding the original meg-CSF		
CC	includes exons II-IV and is thought to terminate in the region between		
CC	amino acid residues 134 - 147. The primary transcript of this gene		
CC	may be cleaved in different ways to yield a family of mRNA's each		
CC	encoding a different MSF protein. Exons V and VI are thought to be		
CC	related to the activity of the factor and are also implicated in the		
CC	stability, folding and processing of the molecule. These exons are		
CC	also thought to play a role in the observed synergy of MSF with other		
CC	cytokines. Exons V - XII are believed to be implicated in the		
CC	processing or folding of the appropriate structure of the resulting		



XX New DNA and related proteins or RNA derived from *M. tuberculosis* -  
 PT used for diagnosis of mycobacterial infections, monitoring  
 PT vaccination and development of anti-mycobacterial agents  
 XX  
 PS Claim 5, Fig 13; 55pp; English.  
 XX  
 CC This novel 74 kDa protein is encoded by an open reading frame of  
 CC a mycobacterium tuberculosis DNA fragment (see A093610) containing  
 CC polymorphic GC-rich sequences. Its amino acid sequence shows  
 CC a high proline content, but there is no homology to any known  
 CC proline-rich antigens of mycobacteria. Novel *M. tuberculosis*  
 CC proteins (see A093610-57) are claimed. These can be produced as  
 CC recombinant proteins, especially in bacterial, yeast, fungal or  
 CC higher eukaryotic host cells, and used for diagnosing tuberculosis  
 CC and other mycobacterial infections in humans or animals. The  
 CC claimed proteins can also be used for epidemiological studies, for  
 CC monitoring vaccination, and for the development of vaccines and  
 CC anti-mycobacterial drugs.  
 CC  
 SQ Sequence 763 AA:

Query Match 9.8%; Score 715; DB 18; Length 763;  
 Best Local Similarity 20.2%; Pred. No. 1.2e-32;

Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

OY 288 PTPKAEETTKGPAITTPKEPTPTTPKEPASTTPKEPTTTIKSAPTTKEPAPTTSAP 347  
 DB 3 PVP-----apralaplpappapapapkskppfpap-----pappcmvlvsaa 46  
 OY 348 TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 407  
 DB 47 PCP--pap-----pappkpkakapfpvpapaparelalplp--pap-----peapre 90  
 OY 408 PAPTPKEPTPTTP-----KEPAPTTPKEPAPTTPK--EPAPTTPK--KPAPTTPKEPAPTTPK 461  
 DB .91 srpalppppppvlpdppeapapvpapnspfpfpfpapkfvpapvp--pvpnspp 148  
 OY 462 EP--APTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 519  
 DB 149 fppfpapalnpap-----paplpanspplpappfpagct--ppaapvpvpaapkskpa 201  
 OY 520 TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 575  
 DB 202 spppppp-----pmpatpmpfpplpvpvpplskcpapppaplpvpvpplp 256  
 OY 576 TTPKKLPTTPTEKLAP-----TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 625  
 DB 257 pvpnkklppap-----appvaavaavlvapcpplpplpnhnpappapavpvpplaplpnshp 313  
 OY 626 ---FKAAAPNTPKEPAP----- 639  
 DB 314 pappapavpvpplaplpisgrpvasvskgstflstfcrcvcsgevlagalnsrpsrsl 373  
 OY 640 --TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 689  
 DB 374 tttcpalpeplpplpplpplpplpplpplpplpplpplpplpplpplpplpplpplppl 431  
 OY 690 LAPPTTKEPSTTSKAPAPTTPKGTAPTTKEPAP-----PTTPKEPAPTTPKGTAPTT 743  
 DB 432 --pfpapkvctfpplapapepk--tvpvlpppscpseknppapppeppespsplp 488  
 OY 744 KEPAPTTP-----KKPAPKELAPTTTKGP--TSTTSKAPAPTTPKEPAPTTPKEPAPTTP 796  
 DB 489 appapmspavrvppspplpappapaprasmpalppapspapalrcplpappapansp 548  
 OY 797 KKPAAPTTPPTPTTSVSTPTTKEPTTIHKSP-----DESTPELSAEPKPA--LENSP 850  
 DB 549 --papapcpfpklls--anpccpvpppapnrvpappapapdelppapppdpvpvansp 604  
 OY 851 KEGVPTTKPAATKPMETTTAKDKTTERDLRTTPTTTTAAPKMTKETATTEKTESKI 910

DB 605 --pappappapsalpfnppa-----ptpeapk-----srpal 637  
 OY 911 TATTQVSTTTQDPTTFEKTITLTKTTTLAPKYTTTKTKTTTTEIMNKPBEAKPKDRATN 970  
 DB 638 paappappapvratlp-----ppappapapn 665  
 OY 971 SKATTPKQKP-----TKAPKPTSTKKPKTMRVRKPKTTPRKMTSTM-----PELNP 1021  
 DB 666 smalpappdpplppllapppappplpmsppappplppaapdpappapplclnqppspplap 725  
 OY 1022 TSRIEAMIQTTTRPQNTPNKSLVEVNPKSEBAGCAEGFT 1061  
 DB 726 vpgaplaplpingrvfarknsl-----gssgdc 756

# RESULT 14

AAV53666  
 ID AAV53666 standard; Protein: 4412 AA.

XX AAV53666;

DT 22-FEB-2000 (first entry)

DE Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.

XX Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;  
 KW bone development; g1/1017427/emb/CAA62189.

XX Unidentified.

XX WO960164-A1.

PD 25-NOV-1999.

XX 14-MAY-1999; 99WO-US11066.

PR 15-MAY-1998; 98US-0085673.

XX (QUAR-) QUARK BIOTECH INC.

PI Elmat P, Mor O, Skallier R, Feinstein E, Faerman A;

XX WPI; 2000-053304/04.

PT Identification of stress induced genes for determining risk and  
 PT preventing, treating or controlling osteoporosis

PS Claim 32; Fig 6A-R; 308pp; English.

CC The present sequence is obtained from a clustal X alignment with  
 CC protein 608. Protein 608 was identified using the method of the invention  
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the  
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected  
 CC to mechanical strain. The specification describes a method for the  
 CC identification of genes responsive to a specific mechanical stress. The  
 CC method comprises applying the mechanical stress to an organism (tissue  
 CC or cells comprising bone cells), isolating the specific cellular  
 CC fractions and extracting mRNA from them, and differentially analysing the  
 CC mRNA in comparison with control samples. The method is used to identify  
 CC genes whose expression is responsive to a specific stress. The identified  
 CC genes are employed in determining risk associated with a physiological or  
 CC disease state. The risk determination methods are used for testing a  
 CC medicament for gene therapy. These medicaments, or genes identified by  
 CC the method of the invention, are used for treating, preventing or  
 CC controlling a physiological or disease state (especially osteoporosis or  
 CC bone density or other factors causing or contributing to osteoporosis or  
 CC its symptoms or other conditions involved in mechanical stress or its  
 CC lack. The methods can also be used for advancing research or studies in  
 CC bone development.

XX Sequence 4412 AA;



```

QY 458 TTKKERA--PTTKESPT--TPKEAPPTTKS-APTTKKPAPT---TTKSAPTTPKEPS 510
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 920 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 979
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 ----PTTKKPAPT--TPKEAPPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTPKPE 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 980 psdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd 1039
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 566 APT---TPKETAPTTPKLLPTTPEKLAPTTPKEKPAPT---TPKEAPTTPKEPTPT--TP 618
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1040 tpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd 1094
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 619 EEPAPTTPKKAAPTTPKEAPPTPKKPA--PTTKKPAPT--TPKETAPTTPKGAAPTTLK 675
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1095 eepilpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd 1154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 676 EPAPTTPKPKPAKELAPTTPKKEPT--STSDKPAPTTPKGAAPTTPKEAPPTPKKPAPTT 734
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1155 eplps--deplpsd--eplpsdeplpsdeplpsdeplpsdeplpsdeplps--deplps- 1208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 735 PKGTAFTTLKEAPPT--TPKKPAKELAPTTPKGAPTSTSDKPAPT---TPKETAPTTPKE 790
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1209 ---deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd--eplpsd 1253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 791 PAPTTPKKAAPT--TPETTPPTSEVSTPTTKKEPTTIHKSPESTPELSAETPKALENS 849
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1254 ptps--deplpsdeplpsdeplpsdeplpsdeplpsdeplps--psdeplps--sdeplpsdeplp 1305
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 850 PKEGCVPTTKTPAAT--KPEMTTAKDKTTERDLRTPEPTTAAPKMTKETATTEKTTE 907
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1306 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd 1361
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 908 SKITATTQVSTTQDTTPFKITLKTTLTTLAPKTTTKITTTTTEIMNKPEETAKPKDR 967
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1362 deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsd 1396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 968 ATNSKAT-----TPKQKPTKAP--KKPTSTKKKTMPTVRKPKKTPTRPKM----- 1012
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1397 sgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg 1453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1013 -----TSMPELNPSTRIAEAMLQTTTRNQTPNSKLVENVKSEDAAGAEGETPHM- 1064
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1454 ylygypdsgfipernitlraaavf-----akl1-----gadesygaqasapsyd 1498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1065 LLRPH-----VEMPE-----VPMDDYLPVRVNOGI----- 1090
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1499 ledthwaawalkfatsgglfkypdgtlkpqnltreafavlnhfltkvayggetmskla 1558
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1091 ---IINPMLSDPTNCGN---KPYDGLTTL-----RNGTLVAFRGHYEWMLSPFSPPS 1137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1559 tldlsnpkfd--cvghwaqeflekltlslylsypdgt-----fkpgn 1600
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1138 PARRITEWAGIPSIDVFTFRCNCEGKTFEKKD--SOYWRFTNDIKD 1182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1601 ylkrsesv-----allnralergplngapklfpdvnesywal-gdlmd 1642
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

XX 09-AUG-2001.
PD
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 27; SEQ ID No 12385; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAI00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosis
XX of diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1325 AA;

```

Query Match 10.8%; Score 794.5; DB 22; Length 1325;  
 Best Local Similarity 26.4%; Pred. No. 7.3e-37;  
 Matches 314; Conservative 109; Mismatches 507; Indels 261; Gaps 46;

```

QY 76 ESFERGREGCDQAQCKKYDKCP-----DYSPFAEVHNPTSPSSKKAPPSGASQT 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51 etheqkrhcntlttshskptdk--ptgnskltdks---stdnheapltseensngqkdpn 105
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 IKSTTRSPRPKPKKKT--KKVIESEITTEYKDNKNRTRKKRPKPPVY---DEAGSGID 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 106 lrr--grsvpadstltthesagknhlpapkskin--czkstgkctvrkcdkgrple 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 NGDFKVTTPDTSTYOHNNKTSPTKITTAKPINRPSLPNSDSQTSKTLTVNKETTVERK 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 k-----smstldkstsahktstfmsngsqtkqkstsfpekltlaasktlytk- 211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 ETTTKKQTSYDGEKERTSA-----KETOSIEKT--SAKDIAFTSKVIAKTPKAEETTTG 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 212 -tgltpesektedsrlltasdklltktnlgetlsanel---tqslaeptelhgtrane 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 PALTTTPKEPTPT---TPKPEASTTPKEPTPTTKSA--PTTKKPAPT-----TTK 344
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 268 mltpspaeplenretanentltspaeplenretanentlapfpaepplenretanentlt 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 SAPTPKKEAPPTTKK--PAPTTPKEAPPTTKKBPAPPTTKSAPTTPKKEAPPTPKKPA 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 lfpaepelngertanentltspaeptelngertane-----ntlpspaepelngertanfnd 383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 PTTTPKKEAPPTTKKEPTP-----TTPKKPAPT--TKKBPAPPTTKKBPAPTKKPAPTTPKEP 455
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 384 ktsssaestehgerltplanentltspaeptlenretanentltspapagptlenretane 443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135132.
PR	24-MAY-1999;	99US-0135353.
PR	25-MAY-1999;	99US-0135629.
PR	27-MAY-1999;	99US-0136021.
PR	28-MAY-1999;	99US-0136392.
PR	01-JUN-1999;	99US-0136782.
PR	03-JUN-1999;	99US-0137222.
PR	04-JUN-1999;	99US-0137528.
PR	07-JUN-1999;	99US-0137502.
PR	08-JUN-1999;	99US-0137724.
PR	10-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	14-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139453.
PR	18-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	29-JUN-1999;	99US-0140823.
PR	30-JUN-1999;	99US-0140981.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	09-JUL-1999;	99US-0142803.
PR	13-JUL-1999;	99US-0142920.
PR	13-JUL-1999;	99US-0142977.
PR	14-JUL-1999;	99US-0143542.
PR	15-JUL-1999;	99US-0143624.
PR	16-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144326.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	20-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144332.
PR	20-JUL-1999;	99US-0144332.
PR	21-JUL-1999;	99US-0144684.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	23-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.

---

PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145293.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	28-JUL-1999;	99US-0145919.
PR	02-AUG-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	03-AUG-1999;	99US-0146389.
PR	04-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147312.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	10-AUG-1999;	99US-0147935.
PR	11-AUG-1999;	99US-0148171.
PR	12-AUG-1999;	99US-0148319.
PR	13-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149368.
PR	18-AUG-1999;	99US-0149175.
PR	20-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	23-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	27-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151438.
PR	07-SEP-1999;	99US-0151930.
PR	10-SEP-1999;	99US-0152363.
PR	13-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154479.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155139.
PR	24-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0156359.
PR	29-SEP-1999;	99US-0156458.
PR	04-OCT-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157117.
PR	06-OCT-1999;	99US-0157755.
PR	07-OCT-1999;	99US-0157865.
PR	08-OCT-1999;	99US-0158029.
PR	12-OCT-1999;	99US-0158232.
PR	13-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.



```

XX 314 Bovine MSF orthologue, superficial zone protein (SZP).
DE
XX
XX Bovine; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
KM superficial zone protein; SZP; MSF orthologue; synovial lubricant;
KM osteoarthritis; joint lubrication; osteopachic; antiarthritic.
XX
OS Bos taurus.
PN WO200107068-A1.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000MO-US20002.
XX
XX 23-JUL-1999; 99US-0145328.
XX 19-JUL-2000; 2000US-0145328.
XX
XX (UYCA-) UNIT CASE WESTERN RESERVE.
XX
XX Warnan ML;
XX
XX WPI: 2001-182721/18.
XX
XX New composition comprising the campodactylly-arthropathy-coxa
PT vara-pericarditis protein in combination with an anesthetic, useful for
PT treating osteoarthritis, or as lubricants of tissue and joints.
XX
XX Example 1; Fig 4; 34pp; English.
XX
XX The invention relates to a method of treating osteoarthritis via the
XX administration of a composition comprising the campodactylly-arthropathy-
XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
XX The composition may further comprise a local anesthetic. The composition
XX of the invention may be administered via intra-articular or intravenous
XX injection. The human CACP protein is identified in the invention as
XX being megakaryocyte stimulating factor (MSF). The gene encoding
XX CACP protein (MSF) is located on chromosome 1q25-31 and mutations in
XX this gene are responsible for the heritable disorder campodactylly-
XX arthropathy-coxa vara-pericarditis, in which patients have synovial
XX hyperplasia without evidence of inflammation. CACP protein (MSF)
XX acts as a synovium lubricant, and can be used to lubricate tissue and
XX joints in the treatment of osteoarthritis. The composition may be
XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
XX loss of range of movement or joint damage). The present sequence
XX represents the bovine orthologue of human MSF, superficial zone
XX protein (SZP).
XX
XX Sequence 472 AA;
SQ

```

```

DB 314 a1slgkyrsrpesvyffkrrgsvqgylykgeptqkclgrpalnysvygetagvrrrrfe 373
OY 1263 RAIGSQTHTRIRIOYSPARLAYODKGVJHANEKYSILMRGJPNVYTSAISIPNIRKPDGQ 1322
DB 374 ra1gspgythlrlhytprrvpygdkgflhnekvksclwrgjpnvvtalsipnlrkpogy 433
OY 1323 DYAFSKDOYYNIDVPSTARAITTRSGQTLISKWYNCP 1361
DB 434 dylalskdqyyndvpsrtaraitrrsgqlslntwncp 472

RESULT 8
AAM24516
ID AAM24516 standard; Protein: 5179 AA.
XX
XX AAM24516;
XX
XX 12-OCT-2001 (first entry)
XX
XX C899P predicted amino acid sequence.
DE
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KM immunogenic; gene therapy; vaccine; colonic cancer.
XX
XX Homo sapiens.
XX
XX WO200149716-A2.
XX
XX 12-JUL-2001.
XX
XX 29-DEC-2000; 2000MO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
XX 10-JAN-2000; 2000US-0480321.
XX 15-FEB-2000; 2000US-0504629.
XX 06-MAR-2000; 2000US-0519444.
XX 19-MAY-2000; 2000US-0575251.
XX 29-JUN-2000; 2000US-0609448.
XX 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 2; Page 446-462; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX PCR, polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(II) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX

```

[illegible]

RESULT	6
AA880041	
ID	AA880041 standard; Protein; 452 AA.
XX	
AC	
AA880041;	
DT	10-APR-1996 (first entry)
XX	
DE	Human megakaryocytopoietin protein.
XX	
KW	Human; megakaryocytopoietin; wheat germ agglutinin; heparin;
KW	megakaryocyte; aplastic anaemia; thrombocytopenia; hematology.
KW	multipotential stem cell.
OS	
XX	Homo sapiens.
XX	
XX	
FT	Key
FT	Location/Qualifiers
FT	Misc-difference 393..396
FT	/note= "unspecified amino acids"
FT	444..446
FT	Misc-difference
FT	/note= "unspecified amino acids"
XX	
XX	
PN	WO9523861-A1.
XX	
PD	08-SEP-1995.
XX	
PF	06-MAR-1995; 95WO-CN00015.
XX	
PR	04-MAR-1994; 94CN-0112066.
XX	
PA	(SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.

XX Gu X, Han Z, Shen Q:  
PI  
XX  
DR WPI: 1995-320576/41.  
N-PSDB: AAT04546.  
XX  
XX New hematopoietic cell growth factor - used for treating  
PI thrombocytopenia and hematocytopenia  
XX  
XX Example; Page 23; 36pp; Chinese.  
XX  
XX This sequence represents the human megakaryocytopoietin (MPO) protein.  
CC This sequence was purified using a carrier which can couple wheat germ  
CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
CC AAR80039 and AAR80040) were used to produce the amplification primers  
CC shown in AAT04544 and AAT04545. The fragments amplified by these primers  
CC can then be used as probes to screen human cDNA libraries for MPO cDNA.  
CC The MPO cDNA can then be inserted into a plasmid which is used to  
CC transform cells to produce MPO. The MPO sequence is capable of promoting  
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes  
CC and stimulating the proliferation of multipotential stem cells. The  
CC factor may be used for treating thrombocytopenia and hematocytopenia.  
CC The purification method can be used to isolate MPO from human urine or  
CC serums of patients with aplastic anaemia, and from animal blood or urine  
CC by radiation exposing the animals to induce aplastic anaemia.  
XX  
XX Sequence 452 AA:  
SQ

Query Match	30.4%;	Score 2228.5;	DB 16;	Length 452;
Best Local Similarity	92.3%;	Pred. No.2.5e-118;		
Matches 419; Conservative	3;	Mismatches 21;	Indels 11;	Gaps 1.
QY	1	MAKTLPIYLILLLSFVILDOVSSODLSSACGAGCGEYSRDATCNCNDYNCONHMECCPFF	60	
Db	1	maktlpiylillllslvfivdgvsqdlssacgfcgsgysrdatcncdyncnhyhmecppdf	60	
QY	61	KRYCIAPLSCKGKCFESFEFGRECDDAOCKKYDKCCDPYESFCAEVHNPTSPSSSKAP	120	
Db	61	kryctaelscckgrcfesfergrecdcaagckkydkccdpysesfcaeinhptspssskap	120	
QY	121	PPGSASQTIKSTTKRSRSPKPNNKKRTKVIESEITEVEKDNKKNRTRKKLPTRPVVDGAG	180	
Db	121	ppgsaasqlsttkrskrpkpnkkrtkvieseeitevkdnnkrrtkklprkpvpvdeag	180	
QY	181	SGLDNGCFKVTTPDTSTTOCHNKYSTSPKITTAIPINRPISLPNSDSKTSLTVNKEET	240	
Db	181	sglndngcfktvtptdsttchnkystspkittaipinrpislpnsdstkslstlvnkett	240	
QY	241	VETKEETTTKNQSTSDCKEKTTSAKENQSILEKTSAKDLAPTISKVLAKPTPKAETTTKGPA	300	
Db	241	vetkeetttknqstsdckekttsakensilektsakdlaptiskvlaprtpkaetltkgpa	300	
QY	301	LTTTKEPLPTTPPKPAASTTPKEPPTTIKASAPTPPKRAPATTTKSAPTPPKEPAPTTTKE	360	
Db	301	ltttekplpttpkpaasttpkep ppttikasaptpkrapatttksaptpkrepatttke	360	
QY	361	PAPTPPKEPAPTTTKEPAPTTTKSAPTTPKREPAPDTTPEKKAPATTPPKEPAPTTTKEEPLPTT	420	
Db	361	paptppkep ap tttkep ap tttksap ttpk rep ap dt p e k k a p at t p k ep ap t t ke e pl p t t	409	
QY	421	PKEPAPTTKEPAPTTTPKEPAPATAKKAPATTPPKE	454	
Db	410	psrlhpptkrepapttkrepapttkshpdrpscxxxxxcqp-----tpkehppl	443	
RESULT 7				
AAB60569				
ID AAB60569 standard; Protein; 472 AA.				
XX AAB60569;				
XX				
DT 27-APR-2001 (first entry)				





```

Db 721 aaptpkpkapkeapltkeptsidskpaclpkytaclpckepapltcpkcapltcpk 780
OY 738 TAPTTKEPAATPKKPAKELAPTTTKGPTSTSDKPAATPKETAPTTKEPAATPK 797
XX 781 taplttkepapltcpkpaclpckepapltcpktsidskpaclpckepapltcpk 840
Db 798 KPAATPEPPTTSEVSTPTTKEPTTIHKSDESEPELSAPPEPKALENSPKEGVP 857
OY 841 kpaclpckepapltcpktsidskpaclpckepapltcpktsidskpaclpckepapltcpk 900
Db 858 TKTPPAATPKPMTTAKDKTERDLRTPETTTAAPKATKETAATTEKTESKTATTT 917
OY 901 tktpaakpkemttakctterdlrttpeettaapkmktetattctesktattctg 960
Db 918 TSTTTODTPEFKITLTKTTLAPKVTTKTTTTEIMNKPEETAKPKDAATNSKATPK 977
OY 961 ttttqcltpkltlcltltlapkvtttkltltteimnkpeetapckdaatnskatcpk 1020
Db 978 POKPTKAPKPTSTKPKPTMPRVKPKTPTPKMTSTMPELNPTSKIAEAMLOTTT 1037
OY 1021 pdkptckapckpckckpckmpvrvkpttpckmtstmpelnsrtaeamlqcltrpn 1080
Db 1038 QTPNSKLTVEVNPKSEDAAGAGETPHMLLRPHVMPPEVTDMOYLPRVPOGIIINP 1097
OY 1081 qtpnsklvevnpksedaagagetrphmlrrphvmppevtmdyldprvpogiiinpm 1140
OY 1098 DETNCKNKPVDGLTTLANGTLVAFRGHYFWMLSPEFSPSPARRITTEWGISPIDT 1157
OY 1141 detnckngkpvdglttllngtlvafrghyfwmllspfpssparlttewgispidt 1200
Db 1158 RCNCGKTEFFEFKDSQYMRFTNDIDKAGYKPIFKGFGGLTQIYAAISTAKYK 1217
OY 1201 rcncgkttffefkdsqywrftndidkagypkpfkgfglqgiyaalstakymwps 1260
OY 1218 FFKRGSGTQYTYKQEPYQKCGRRPAINPYVGMTOVRRRRPRAIGPOTITRI 1277
OY 1261 ffkrgsgtqylykqepvqkcpgrrrpalnpyvgmtgyrrrrferralgpsq 1320
OY 1278 SPARLAYOQKGYLHNEVKSILMRGLPNVVTSAISLPNIRKPDGYDYAFSK 1337
Db 1321 sparlavgdkylylhnvksllwrglpnvvtasalslpnrkpdgydyafskdgy 1380
OY 1338 PSRTARATITRSQGLSKWYNCP 1361
Db 1381 psrtaratitrsqglskwvynp 1404

RESULT 4
ID AAM24322 standard; Protein: 1299 AA.
XX AAM24322;
AC AAM24322;
XX 12-OCT-2001 (first entry)
DT 12-OCT-2001 (first entry)
XX Human EST encoded protein SEQ ID NO: 1847.
DE Human
XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX Homo sapiens.
OS Homo sapiens.
XX MO200154477-A2.
PN MO200154477-A2.
XX 02-AUG-2001.
PD 02-AUG-2001.
XX 25-JAN-2001; 2001MO-US02687.
PF 25-JAN-2001; 2000US-0491404.
XX 25-JAN-2000; 2000US-0617746.
PR 17-JUL-2000; 2000US-0617746.

```

```

PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-063870.
XX (HYSE-) HYSEQ INC.
PA Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX Cao Y, Dimaic RA, Zhang J, Werhman T;
XX WPI: 2001-476164/51.
XX N-PSDB: AAH98981.
DR Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use
XX Claim 20; Page 1198-1201; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a
XX protein of the invention.
SQ Sequence 1299 AA;
Query Match 91.7%; Score 6715.5; DB 22; Length 1299;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1252; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
OY 1 MAMKTLPTLYLLLSVFVIOQVSSODLSSCAGCGEGYRDATCNDYCMQHYMECCP 60
Db 1 maktlptlyllllsvflqvsgslsscagcgysrdatcndyqhyemccpdt 60
OY 61 KRVCTAELSCGKCFSEFERGRCDDAOCKKYDKCCCPYDFSCAVHNPSTSPSSKAP 120
Db 61 krvctaelscgkcfsefergrecddaockkydkccpdydfscavhnpstspssk 120
OY 121 PPSASOTIKSTTKRSPKPPNKKTKVIESEITE----- 156
Db 121 ppsasotiksttkrskppnkktkvieseeiteehsvsengessssssssstliw 180
OY 157 -----VKDNKKNTKKKPPKPPVVDASGLDNGFKTTPDTST 197
Db 181 kiksksnaaurelqkklkykdnkntkkkppkppvvdasgldngfkvtptdst 240
OY 198 TOHKKVSPKITTAKPINPPSLPPNSDTSKETSITVNRKETTVERKETTNNKOST 257
Db 241 tqhmkvscspkittakpinpsslppnsdtsketsitvnrkettvetkettlnkq 300
OY 258 KEKTSAKETOSIEKTSADLAPTSKYLAKTPKAEITTTGCPALTTPKEPTTPK 317
Db 301 kektsaketqsiektsadlaptskylvlaktpkaeitttgcpalttpepttpk 360
OY 318 TTPKEPTPTTITKSAPTTKEPAATPKKPAATPKKPAATPKKPAATPKKPAAT 377
Db 361 ttpkeptpttitsapttkepaatpkpkpaatpkpkpaatpkpkpaatpkpkpa 420
OY 378 APTTKSAPTPKKEAPPTPKKPAATPKKPAATPKKPAATPKKPAATPKKPAAT 437
Db 421 apttksaptpkpeaptpkpkpaatpkpkpaatpkpkpaatpkpkpaatpkpk 480
OY 438 EPAPTAPKPPAPTTTKKEAPPTPKKPAATPKKPAATPKKPAATPKKPAAT 497
Db 481 epaptapkpaptttkkeaptpkpkpaatpkpkpaatpkpkpaatpkpkpaat 540
OY 498 TTKSAPTPKKESPTTKEPAATPKKPAATPKKPAATPKKPAATPKKPAAT 557
Db 541 tksaptpkkespttkpeaptpkpkpaatpkpkpaatpkpkpaatpkpkpaat 600
OY 558 APTAPKEPAATPKKETAATPKKLTPTTPEKLAPTTPEKPAATTPPELAPTT 617

```

QY 978 POKPKAKPKTSTKKPKTPMRVRKPKTPTPKMTSTMBELNPTSRITAEAMLOTTTRPN 1037  
DB 1021 PqKpLkApKkKpLkKpKpLmPrVrkKpLtpPrKmtStMpeLnpLstRiaeAmLqTtRpn 1080  
QY 1038 QTPSKLVEVNPKEDEGAGEGEPHMLLRPHVPMPEVTDMDYLPRVPMOGIINPMIs 1097  
DB 1081 qtpmsklvevnpkdedagaegeclpmlLrphvfmpcvrpdmyLprvpqgIilnplms 1140  
QY 1098 DETNICKGKPVGDLTTLRNGTLVAFRGHPMMLSPFSPSPARIRTEVWGISPIDVT 1157  
DB 1141 deLtnCngKpVdglTtLrNgTlVafRghyFmLspfsPsparrtIevwIspIdtVt 1200  
QY 1158 RCMCEGTEFFEKDSQVYRFPNDKDGAPYPRIRKGFGLGTGIVAAIStAKKXNMPESY 1217  
DB 1201 rcmcegtEtfFkdsgYwrfndkldagYpKpLkfgIgtqIvaalStakYknwPesy 1260  
QY 1218 FFRGSGSIQYIYKQEPVQKCPGRPALNVPYGEHTOYRRRFEALISQTHIRIOY 1277  
DB 1261 ffrgsgsiqYIYkgErvKcpgrPalnYpygEmtYrrrrfEraIaIpsqchLtrIqY 1320  
QY 1278 SPARLAYOOKGVLAHNEKYSILMRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYXNDY 1337  
DB 1321 sparIayqdgYvLhnevKvslwrgIpnvvtSaIsipnIrKpdgydyafskdqYyndv 1380  
QY 1338 PSRTARAITTRSGOTLSKVMYVNC 1361  
DB 1381 psrtaraittrsgqtIskVynncp 1404

RESULT 3  
AAB29773  
ID AAB29773 standard; Protein: 1404 AA.  
AC AAB29773;  
XX  
DT 28-FEB-2001 (first entry)  
XX  
DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO.1.  
XX  
KW Human MSF: megakaryocyte stimulating factor; tribonectin;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KW friction coefficient reduction; gene therapy; antiarthritic;  
KW osteoponthic.  
XX  
OS Homo sapiens.  
XX  
PN MO200064930-A2.  
XX  
PD 02-NOV-2000.  
XX  
PF 24-APR-2000; 2000MO-US10953.  
XX  
PR 23-APR-1999; 99US-0298970.  
XX  
PA (RHOD-) RHODE ISLAND HOSPITAL LIFEESPAN PARTNER.  
XX  
PI Jay GD;  
XX  
DR WPI: 2001-024673/03.  
XX  
DR N-PSDB; AAC81498.  
XX  
PT Novel tribonectin polypeptide useful as lubricant for treating  
PT osteoarthritis, comprises O-linked lubricating moiety  
XX  
PS Claim 3; Page 7; 47pp; English.  
XX  
CC The invention relates to a human tribonectin which is a product of  
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonectin has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76  
CC repeats of a motif having at least 50% identity to the sequence KRPAPPT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a

CC human MSF-derived tribonectin; a biocompatible composition comprising a  
CC human tribonectin for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC or trauma of a mammal, where the injury is caused by a surgical insertion  
CC of a particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonectin may be used in gene therapy. The present sequence represents  
CC human MSF.  
XX  
SQ Sequence 1404 AA:

Query Match 99.6%; Score 7291.5; DB 22; Length 1404;  
Best Local Similarity 96.9%; Pred. No. 0;  
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQOVSQDLSSCAGRCBGYSKDATCNDYNCQHYMECCPDF 60  
DB 1 mamktlpiYlllsvfIqvsqDlssCagrcBgysrdatcncdYncqhyMecpDf 60  
QY 61 KRVTAELSCGRCFESFERGECDDAOCCKYDKCCPDYECFCAVHNTPSPSSKAP 120  
DB 61 krvtaelScgrcfesfergEcddAocckYdkccpdyEcfcavhNtpssSkap 120  
QY 121 PPSGASQTISTTKRSPKPPNKKKKVIESEITE----- 156  
DB 121 ppsgaSqtIsttkrSpkppNkkkkvIesEite----- 180  
QY 157 -----VKDNKKNRKKKPPKPPVYDVGSGLDNDFKVTPTDST 197  
DB 181 kksksksaanrelqkklvkdakknrtkkpdkpvydvgsgldnDfKvtptDst 240  
QY 198 TQNNKSTSKITTYAKPINRPSLPNDSKETSJLVNKETVETTKTNNKOSTG 257  
DB 241 tqnnkStskIttyakPinrpsLpndsKetsJlvnKetvEtTKtnnKostg 300  
QY 258 KEKTSAKETOSIEKTSANDIAPTSKVLAKPPPKATETTKGALTYTPKEPTTREPAS 317  
DB 301 kektsaketosIektsanDIAptskvLakppPkAtetTKgAltytpKepttRepas 360  
QY 318 TTPKEPTPTTIKSAPTTPEKAPATTTKASAPTTPEKAPATTKEPAPTTPKEP 377  
DB 361 ttpkeptptTIksapTTpEkapATTtKasapTTpEkapATTKEpAPTTPKEp 420  
QY 378 APTTTSAPTTKEKAPATTPKKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 437  
DB 421 aptttsapTTkEkapATTpKkapATTpEkapATTpEkapATTpEkapATTpEk 480  
QY 438 EPAPTAAPKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEKAP 497  
DB 481 epaptaAPkApATTpEkapATTpEkapATTpEkapATTpEkapATTpEkap 540  
QY 498 TTKSAPTTPEKESPTTTEKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 557  
DB 541 ttksapTTpEkEsptTTtEkapATTpEkapATTpEkapATTpEkapATTpEk 600  
QY 558 APTAPKEPAPTTPKEPATTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPE 617  
DB 601 aptapkePAPTtpKEpATTpEkapATTpEkapATTpEkapATTpEkapATTp 660  
QY 618 BEEPAPTTPKAAAPNTPEKAPATTPEKAPATTPEKAPATTPEKAPATTPEK 677  
DB 661 beepapTTpkAAapNTpEkapATTpEkapATTpEkapATTpEkapATTpEk 720  
QY 678 APTTPKKAPKELAPTTTKEPTSTSDKPAATTTKGAPATTPEKAPATTPEKAP 737

QY 1218 EFRKGGSTIOOYITKOPVOKCPGRPALNYPYGGEMTOVRRRRFRRAIGPSQTHIRIQY 1277  
 DB 1261 ffrkrgssiqylykgepvqkcpripalpyvgentgyrrrrfrfcaiaqpsqthiririy 1320  
 QY 1278 SPARLAVODKGVJHNEVKVSIIMRGLPNVVTSAISLPNTRKPDGDYAFSSQDYINIDV 1337  
 DB 1321 sparlavodkgvjlhnevksilwrglpnvvtalsipntrkpdgydyafsskdyynldv 1380  
 QY 1338 PSRTAPATITRRSGOTLSKMYNCP 1361  
 DB 1381 psrtaraitrsgqlskmyncp 1404  
 RESULT 2  
 AAB60568 standard; Protein; 1404 AA.  
 AAB60568:  
 27-APR-2001 (first entry)  
 Human megakaryocyte stimulating factor (MSF, CACP).  
 Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;  
 MSF; megakaryocyte stimulating factor; synovial lubricant;  
 chromosome 1925-31; osteoarthritis; joint lubrication; osteopathic;  
 antiarthritic.  
 Homo sapiens.  
 WO20107068-A1.  
 01-FEB-2001.  
 21-JUL-2000; 2000WO-US20002.  
 23-JUL-1999; 99US-0145328.  
 19-JUL-2000; 2000US-0145328.  
 (UYCA-) UNIV CASE WESTERN RESERVE.  
 Warman ML.  
 WPI: 2001-182721/18.  
 New composition comprising the campodactylly-arthropathy-coxa  
 vara-pericarditis protein in combination with an anesthetic, useful for  
 treating osteoarthritis, or as lubricants of tissue and joints  
 Example 1; Page -: 34pp; English.  
 The invention relates to a method of treating osteoarthritis via the  
 administration of a composition comprising the campodactylly-arthropathy-  
 coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 The composition may further comprise a local anesthetic. The composition  
 of the invention may be administered via intra-articular or intravenous  
 injection. The human CACP protein is identified in the encoding as  
 being megakaryocyte stimulating factor (MSF). The gene encoding  
 CACP protein (MSF) is located on chromosome 1925-31, and mutations in  
 this gene are responsible for the heritable disorder campodactylly-  
 arthropathy-coxa vara-pericarditis, in which patients have synovial  
 hyperplasia without evidence of inflammation. CACP protein (MSF)  
 acts as a synovium lubricant, and can be used to lubricate tissue and  
 joints in the treatment of osteoarthritis (e.g., joint pain,  
 applied to reduce the symptoms of osteoarthritis). The present sequence  
 is of range of movement or joint damage). The present sequence  
 represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U070316).  
 Sequence 1404 AA:  
 SQ

Query Match 99.6%: Score 7291.5; DB 22; Length 1404;  
 Best Local Similarity 96.9%: Pred. No. 0;  
 Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
 QY 1 MAMKTLPIYLLLLSVFIQOVSSODLSSCAGRCGEYSRDATNCNDYNCOHMECCPDF 60  
 DB 1 mawktlpiyllllsvfifqgvssddlsscagrcgeysrdatncndynohmccpdf 60  
 QY 61 KRVTAEISCGRCPESEEREGEDCDQCKKYDKCQDYPSCAEVHNPTSPSSKKAP 120  
 DB 61 kvctaelscgrcpeeserergedcdqckkydkcqdypscavhnpstpssskkap 120  
 QY 121 PPSGASQITKSTTKRSPKPPNKKTKKVIIESEILTE----- 156  
 DB 121 ppsgasqltksttkrspkppnkktkkviieseeltehsaengessssssstliw 180  
 QY 157 -----VKDNKNRTKKKPPKPPVYDAGSLGNDGDFKVTTPDST 197  
 DB 157 -----vkdnknrtkkkppkppvydagslgndgdfkvttpdst 240  
 QY 198 TQHNKYSTSPKITTAKPIINRPSLDPNSDTSKETSITVNEKETTNTKOTSTDG 257  
 DB 198 tqhnykstspkittakpiinrpsldpnsdtsketsitvneketntkqstldg 300  
 QY 241 TQHNKYSTSPKITTAKPIINRPSLDPNSDTSKETSITVNEKETTNTKOTSTDG 300  
 DB 241 tqhnykstspkittakpiinrpsldpnsdtsketsitvneketntkqstldg 300  
 QY 258 KEKTSAKETOSIEKTSANDLAPTSKVLAKPPKAEITTKGALTPPKPPPTTKPEPAS 317  
 DB 301 kektsaketosiektsandlaptskvlakppkaeittkkgaltppkpppttkpepas 360  
 QY 318 TTPKEPTPTTIKAPTPKEPATTTKSAPTTPKEPATTPKEPATTPKEPATTPKEP 377  
 DB 361 ttpkeptpttikaptpkepatttksaptpkepattpkepattpkepattpkepattpk 420  
 QY 378 APPTTSAPTPPEPATTPPKKAPATTPPEPATTPKEPTPTTPKEPATTPKEPATTPK 437  
 DB 421 apttksaptpkepattpkpkaptpkepattpkepattpkepattpkepattpk 480  
 QY 438 EPAPTAPEKAPATTPPEPATTPPKKAPATTPPEPATTPKEPATTPKEPATTPKEPAT 497  
 DB 481 epaptapekaptppepattpkpkaptppepattpkepattpkepattpkepattpk 540  
 QY 498 TTKSAPTPPEPATTPPEPATTPPKKAPATTPPEPATTPKEPATTPKEPATTPKEPAT 557  
 DB 541 tksaptppepattppepattpkpkaptppepattpkepattpkepattpkepattpk 600  
 QY 558 APAPKAPAPATTPPEPATTPPKKAPATTPPEPATTPKEPATTPPEPATTPPEPATTP 617  
 DB 601 apapkapaptppepattpkpkaptppepattpkepattppepattppepattpk 660  
 QY 618 PEPAPATTPKAAAPNTPEPATTPPEPATTPPEPATTPKEPATTPKEPATTPKEPATTP 677  
 DB 661 peepaptpkkaapntpepattppepattppepattpkepattpkepattpkepattpk 720  
 QY 678 APPTPKKAPAPKELAPTTKEPTSTTSKPAVTPPKGAPTTPKEPATTPKEPATTPPKG 737  
 DB 721 aptpkkapapkelapttkeptsttskpaavtppkgapttpkepattpkepattpkg 780  
 QY 738 TAPPTLKAPATTPPKKAPKELAPTTKGPSTSDPAPATTPPEPATTPPEPATTPPK 797  
 DB 781 tapptlkapattpkkapkelapttkgsstsdpapattppepattppepattpk 840  
 QY 798 KPAPTPPEPTTPPEPATTPPEPATTPPKKAPATTPPEPATTPPEPATTPPEPATTP 857  
 DB 841 kpaptppepttppepattppepattpkkapattppepattppepattppepattpk 900  
 QY 858 TKTPATPEMTTAKDKTTERDLRTPEETTTAAPKMTKEATTTKESKIRATTTQV 917  
 DB 901 tktpatpemttaakdktterdlrtpeettttaapkmtkeatttkeskiratttqv 960  
 QY 918 TSTTODPTPEKTTTLKATTTLAPKVTYTTTKKTTTTEIMNKPEEAKPKDRATNSKATTPK 977  
 DB 961 tsttqdpktttlkatttlapkvtytttkktttteimnkpeeakpkdratnkattpk 1020

FT Region 1266..1331  
 FT /label= Exon\_X  
 FT 1331..1373  
 FT /label= Exon\_XI  
 FT 1373..1404  
 FT /label= Exon\_XII  
 XX MO9213075-A.  
 PN  
 PD 06-AUG-1992.  
 XX  
 XX 17-JAN-1992: 92WO-US00433.  
 XX  
 XX 18-JAN-1991: 91US-0643502.  
 PR 10-SEP-1991: 91US-0757022.  
 PA (GENY ) GENETICS INST INC.  
 XX  
 PI Clark SC, Geeser TG, Hewick RM, Jacobs K, Turner K;  
 DR WPI: 1992-284660/34.  
 DR N-PSDB: AAQ27223.  
 XX  
 XX New human mega-karyocyte stimulating factors - for treating  
 PT immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.  
 XX  
 PS Claim 1, 2 and 3: Fig 1: 87pp; English.  
 CC The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine proceeding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC primary transcript of this gene may be cleaved in different ways to  
 CC yield a family of mRNAs each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor, ie. one or more of these exons may contain  
 CC sequences which direct proteolytic cleavage, adhesion, organisation of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 XX  
 XX Sequence 1404 AA:  
 QY 1 MAMKTLPIYLLLSVFIQOVSSODLSSCAGRCGEGYSNDATCNDYNOAHMECCPFP 60  
 Db 1 mawktlpiyllllsvfivlqvgasqdlsscagrcggyrdatcndcynqymeccpf 60  
 QY 61 KRYCTAELSCGRGCFSEFGRGRCDDAQCCKYDKCCPDYEEFCABVHNPTSPSSKKAP 120  
 Db 61 krcyteaelscgrgcfsefgrgrrcddaqckkydkccpdyeefcabvhnptspsskkap 120  
 QY 121 PPSGASQTIKSTKRSRPPKPKKTKKVTESFEITE----- 156  
 Db 121 ppsgasqtkstktrspkppnkktkkvteelhsvengeassssssssssstliw 180

QY 157 -----VKDNKKRKKRKKRPPKPPVVDAGSLDNGDKVTPDTST 197  
 Db 161 kksksksaanrejqkklkvdknknrkkktpkppvvdagsgldngdfkvtctpstc 240  
 QY 198 TOHNVSTSPKITTAKPKNPRLPNSDTSKETSLVYANKETVETKETTNNKOTSDG 257  
 Db 241 tqhnkvsctspklttakpknprslpnsdstsketslvynekvtetkettntkqscdg 300  
 QY 258 KEKTSAKETOSIEKTSKDLAPTSKVLAKPTPAETTTKGPALTTKPEPTTPPKEPAS 317  
 Db 301 kektsakeitstlektsskdlaptskvlakpdpkaetttkypalttkpepttkpepas 360  
 QY 318 TTPKEPTPTTKSAPTTPKKEPAPTTKSAPTTPKKEPAPTTKKEPAPTTTKEP 377  
 Db 361 ttpkeptpttksapttpkkepapttkpsapttkkepapttkkepapttkkep 420  
 QY 378 APTTTKSAPTTPKKPAPTTPKKPAPTTPKKEPAPTTKPEPTTPPKKEPAPTTK 437  
 Db 421 aptttksapttkpapttkpkpapttkpapttkpapttkpapttkpapttkpapttk 480  
 QY 438 EPAPTAEPKPAPTTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKSAPTTK 497  
 Db 481 epaptakpapttkpapttkpapttkpapttkpapttkpapttkpapttkpapttk 540  
 QY 498 TTKSAPTTKPEPSTTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 557  
 Db 541 tksapttkpesspttkkepapttkkepapttkkepapttkkepapttkkepapttk 600  
 QY 558 APTAPKPAPTTPKKEPAPTTPKKLTPTTPPKLAPTTPEKAPPTTPBELAPTTPEEP 617  
 Db 601 aptapkepapttkpapttkpapttkpapttkpapttkpapttkpapttkpapttk 660  
 QY 618 PEEBAPTTKPAAPNTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAP 677  
 Db 661 peebapttkpaapntpkkepapttkkepapttkkepapttkkepapttkkepapttk 720  
 QY 678 APTPKKPAKELAPTTTKEPTSTSDKAPPTPKGATPTTKEPAPTTKKEPAPTTPK 737  
 Db 721 aptpkpkapelaptttkeptstsdkapptpkgatpttkkepapttkkepapttkp 780  
 QY 738 TAPTTKEPAPTTPKKPAKELAPTTTKEPTSTSDKAPPTPKGATPTTKEPAPTTPK 797  
 Db 781 tapttkepapttkpkpapelaptttkeptstsdkapptpkgatpttkkepapttkp 840  
 QY 798 KPAPTPPEPPTTSVSTSTTKEPTTHKSPDESPELSAEPKALENSPEKPEGYPT 857  
 Db 841 kpaptpppppttsvststtkeptthkspdespeelsaepkalaenspekpegypt 900  
 QY 858 TKTPAATKPEMTTAKDKTTERDLRTPTTAAKMKETATTEKTTESKITTATTOY 917  
 Db 901 tktpaatkpemttakdkterdlrtpttAAKMKETATTEKTTESKITTATTOY 960  
 QY 918 TSTTODTTPPEKITTLLKTTTTLAPKVTTKTITTTTINKKPEETAAPKDRATNSKAT 977  
 Db 961 tsttoddtppekittllktttllapkvttkktittlelnmkpeetakpdratnskat 1020  
 QY 978 POKPTAPKPKPSTKPKPTMPVRKPKTTPPKKTTSTMPELNPTSRJAEMLOTTTRPN 1037  
 Db 1021 pokptapkpkpstpkpmpvrkpktpptpkktstmpelnptsrjaemlottrpn 1080  
 QY 1038 QTPNSKLVENVKSDAGAGETPHMLLRPHVEPVTTPMDVLPRVNGGIIITNMLS 1097  
 Db 1081 qtpnsklvenvksdagagetphmlrrphvepvttpmdvlprvnggiiitnmls 1140  
 QY 1098 DETNCKNKPVDGLTTLNGLTVAFRGHYFWMLSPFSPSPARRITEVWGPSIDTVEFT 1157  
 Db 1141 detncknkpvdglttlngltvafrghyfwmllspfspsparritevwgpsidvtvc 1200  
 QY 1158 RCNCEGKFFFKDSQYMFETNDIDAGYPRIFKGFGLGQIYAALSTAKKXWMPESVY 1217  
 Db 1201 rcncegkffffkdsqyrfndidagypriifkgfglgqiyaaalstakxwnmpesvy 1260

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:13 ; Search time 138.84 seconds  
(without alignments)  
726.115 Million cell updates/sec

Title: AA7  
Perfect score: 7323  
Sequence: 1 MAWKTLPIYLILLSVFVIO.....AAITRSGQTLSKWVWNC P 1361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDS8/gcgdata/geneseq/AA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/AA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/AA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/AA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/AA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/AA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/AA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/AA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/AA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/AA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/AA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/AA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/AA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/AA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/AA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/AA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/AA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/AA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/AA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/AA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/AA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7291.5	99.6	1404	AA26049	MSF precursor. Sy
2	7291.5	99.6	1404	AA26056	Human megakaryocyt
3	7291.5	99.6	1404	AA26073	Human megakaryocyt
4	6715.5	91.7	1299	AA24322	Human EST encoded
5	3484	47.6	902	AA29778	Human MSF-derived
6	2228.5	30.4	452	AA80041	Human megakaryocyt
7	1707.5	23.3	472	AA26056	Bovine MSF ortholo
8	1194	16.3	5179	AA24516	C899P predicted am
9	950	13.0	763	AA23942	Arabidopsis thalia
10	794.5	10.8	1325	AA43106	C. thermophilum OI
11				AA203645	Peptide #2327 enco

12	756	10.3	182	12	AA10872	Protein encoded by
13	715	9.8	763	18	AA31852	Mycobacterium tube
14	643	8.8	4412	21	AA53666	Sequence g1/01742
15	613	8.4	111	13	AA26050	MSF-K130. Synthet
16	612	8.4	572	18	AA31855	Mycobacterium tube
17	552.5	7.5	844	7	AA60570	Sequence of the fa
18	542	7.4	807	21	AA54467	Amino acid sequenc
19	520.5	7.1	788	21	AA54466	Amino acid sequenc
20	510.5	7.0	1837	21	AA11726	Cryptosporidium pa
21	506.5	6.9	744	9	AA2975	Bioadhesive precu
22	496.5	6.8	1721	19	AA48299	Cryptosporidium pa
23	496.5	6.8	1721	21	AA11727	Portion of Cryptos
24	486.5	6.8	2972	22	AA50362	Human ORFX ORF95
25	489.5	6.7	3118	22	AA50362	Human SRCAP. Homo
26	489.5	6.7	826	13	AA26042	Human SRCAP. Homo
27	488.5	6.7	617	22	AA16458	P. yoelii SSP2 ant
28	488	6.7	617	22	AA16458	Peptide #2892 enco
29	488	6.7	957	22	AA59288	Human MUC11 polype
30	488	6.7	957	22	AA59288	C900P predicted am
31	488	6.7	957	22	AA24513	Human protein sequ
32	485	6.6	1127	22	AA89541	Bioadhesive precu
33	476.5	6.5	652	9	AA2974	Peptide #1317 enco
34	467.5	6.4	511	22	AA14883	Peptide #1319 enco
35	467.5	6.4	511	22	AA27312	Peptide #1289 enco
36	467.5	6.4	511	22	AA2607	Human 07C27 gene
37	451	6.2	2819	22	AA35408	PRP 378. Trificum
38	450.5	6.2	378	12	AA14160	Peptide #967 enco
39	449.5	6.1	2665	22	AA14533	Peptide #987 enco
40	449.5	6.1	2665	22	AA26950	Peptide #941 enco
41	449.5	6.1	2665	22	AA26259	Human ORFX ORF2255
42	449.5	6.1	3266	12	AA42491	Human ORFX ORF2255
43	446.5	6.1	378	12	AA14162	PRP encoded by clo
44	446.5	6.1	750	20	AA505477	C. albicans Rb11 p
45	445.5	6.1	1012	20	AA17406	Human atrophin-1 r

ALIGNMENTS

AA26049	1	AA26049	standard; Protein; 1404 AA.
XX	XX	AA26049:	
AC	XX	02-FEB-1993	(first entry)
XX	XX		
DE	XX	MSF precursor.	
XX	XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;	
KW	XX	stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	XX	Synthetic.	
XX	XX	Location/Qualifiers	
FT	FT	1..26	/label= Exon_I
FT	FT	26..67	/label= Exon_II
FT	FT	67..107	/label= Exon_III
FT	FT	107..157	/label= Exon_IV
FT	FT	157..200	/label= Exon_V
FT	FT	200..1141	/label= Exon_VI
FT	FT	1411..1166	/label= Exon_VII
FT	FT	1166..1212	/label= Exon_VIII
FT	FT	1212..1266	/label= Exon_IX
FT	FT		

Protein encoded by  
Mycobacterium tube  
Sequence g1/01742  
MSF-K130. Synthet  
Mycobacterium tube  
Sequence of the fa  
Amino acid sequenc  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precu  
Cryptosporidium pa  
Portion of Cryptos  
Human ORFX ORF95  
Human SRCAP. Homo  
P. yoelii SSP2 ant  
Peptide #2892 enco  
Human MUC11 polype  
C900P predicted am  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1319 enco  
Peptide #1289 enco  
Human 07C27 gene  
PRP 378. Trificum  
Peptide #967 enco  
Peptide #987 enco  
Peptide #941 enco  
Human ORFX ORF2255  
PRP encoded by clo  
C. albicans Rb11 p  
Human atrophin-1 r

p. 1-3





XX WPI: 1986-094065/14.  
 DR N-PSDB: AAN60473.  
 XX DNA coding for Plasmodium falciparum antigens - expressing  
 PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P  
 XX falciparum  
 XX  
 PS Disclosure: Fig 7: 55pp: English.  
 CC The inventors claim a novel DNA molecule which comprises a  
 CC nucleotide sequence corresp. to all or a portion of the base  
 CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have  
 CC antigenicity suitable for providing protective immunity against  
 CC Plasmodium falciparum malarial infections.  
 XX  
 SQ Sequence 844 AA;

Query Match 7.88; Score 547.5; DB 7; Length 844;  
 Best Local Similarity 26.6%; Pred. No. 6.5e-24;  
 Matches 248; Conservative 118; Mismatches 359; Indels 209; Gaps 49;

QY 99 DYSCAENVKDNKKRTKPKPVVDKSGLDNGDFKVTTPDSTQHNKVSPPK 158  
 DB :  
 DB 16 EYSSLSNENPQBATLTLTPDQ-LSAALNHPN-ELNLETSLTLNENENP 72  
 QY 159 ITTAKPINRPSLPNS-----DTSKETSLVANKETVETKETTNTKOTSTOCKE 209  
 DB :  
 DB 73 VPSAIS-NTLDTLCPYQLIIISLVNFCRKSSQLYNK--NFEKFNLSAVSSNATQGE 129  
 QY 210 KTTSAKETQSIETKSANDLAPTSVLAKPTPKAETTTGPAITTPKEPT---PTTPKEP 265  
 DB :  
 DB 130 NANGNKELNEV-KESSQCPV-----LPQELVTIQ--AAALPQELVEIQEVLIEEP 179  
 QY 266 ASTTPKEPTPTTISAPTPKEPAITTTKSAPTPKEPAITTPKEPAITTPKEPA-----P 321  
 DB :  
 DB 180 VLT-GEVLT---GEPLTVQEP--VTVGEVLTQGE--PVTQGE--PVTQGEVLTQGE 227  
 QY 322 TTTKEPAITTTKSAPTPKEPAITTPKPKPAITTPKEPAITTPKEPAITTPKEPAITTPK 381  
 DB :  
 DB 228 VTVGEPT---VTSQGEVLTQGE-----PVTQGE--PVTQGEVLTQGEVLTQGE 269  
 QY 382 A-----PTTPKEPAITTPKPKPAITTPKEPAITTPKEPAITTPKEPAITTPKEP 429  
 DB :  
 DB 270 VLTQEPVLTIEEPVLT---GEVLTIEE--PVTQGEVLTIEEPVLTQGEVLTQGE 323  
 QY 430 APTTTKSAPITTPKEPAITTP-----TTKSAPITTPKEPSPTTPKEPA----- 468  
 DB :  
 DB 324 ---VLTQEPVLTQGEVLTIEEHLDEKKSQEGDNLSISLSEETEEKSHLKKKSSWLKFG 380  
 QY 469 -----PTTPKEPAITTPKPKPAITTPKEPAITTPKPKPAITTPKPKPAITTPK 521  
 DB :  
 DB 381 GNKNDKSKNEKPSLESVKNADQKEQPLDSQISVNAQSVLTQGEPTQGE-PTTQGE 439  
 QY 522 ETAVTTPKTLPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPK 581  
 DB :  
 DB 440 LTA-----TQGE--PTTQGE--TVLEGEPTTQGE-----TVTAQEP 470  
 QY 582 NTPKEPA-----PTTPKE-----PAITTPKEPAITTPKETAITTPKGTAPITLKEPAITTPK 633  
 DB :  
 DB 471 ILTQEPVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGE 530  
 QY 634 KPAKELAPITTPKEPTSTSDKRAPITTPKGTAPITTPKEPAITTPKEPAITTPKGTAPITL 693  
 DB :  
 DB 531 KSAWLKIGYGNKNDKSKNEKPSLESVKNADQKEQPLDSQISVNAQSVLTQGEPTAT 590  
 QY 694 KEBAITTPKPKAPKELAPITTPKGTSTSDKRAPITTPKETAITTPKETAITTPKETAITTPK 753  
 DB :  
 DB 591 QEP-PTQGEVLTQGE--PTTQGE--ETVLEGEPTTQGE--VTAQEPVLT---QEPVLT 637  
 QY 754 PETTPPTTSVSTPTTPKEPTTHKSPDE-----STPELSAETTPKA----- 795  
 :

DB 638 AQ-EPTVLTQEL---IATQEPVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGE 693  
 QY 796 -----LENSPKKEPGVPTTPKPAITTPKEMTTTAKDITTPKEPTTPKAPK 843  
 DB 694 LKFGYGNKNDKSKNEKPSLESVKNADQKEQPLDSQISVNAQSVLTQGEPTATQGE 753  
 QY 844 MTKETATITTEKTTESKITATTTQVTSSTTQDTTPPK--ITTLKTTTTLAPKVTT-----TK 896  
 DB :  
 DB 754 LTQELTATQEPVLTQGE--TVLEGEPTTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGEVLTQGE 810  
 QY 897 KTTTTEIMNKPPEETAKPKDRATNSKATTPKPK 930  
 DB :  
 DB 811 EPSTTQGE--HDEKKSQEGDNLSISLSEETEEK 842

Search completed: April 26, 2002, 16:25:54  
 Job time: 544 sec







```

QY 408 TTKKEPA--PTTKERSPT--TPKEPAPTTKS--APTTKKEAPT---TTKSAPTTPKES 460
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 920 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 979
QY 461 ----PTTKKEAPT--TPKEBAPTTPKKAPPTTPKEBAPTTPKEBAPTTPKKAPPTAPKEP 515
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 980 psdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1039
QY 516 APT---TPKEBAPTTPKKAPPTTPKEBAPTTPKEBAPT---TPKEBAPTTPKEBAPTTP 568
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1040 tpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1094
QY 569 EEPBAPTTPKAAPNTTPKEBAPTTPKEBAPTTPKEBAPT--TPKEBAPTTPKEBAPTTPKA 625
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1095 eepipidpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1154
QY 626 EBPAPTTPKKAPKEBAPTTPTKKEPT--STSDKAPPTTPKGTAPPTTPKEBAPTTPKEBAPT 684
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1155 eptps--deplpsd--eptpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1208
QY 685 PKGTAPPTLKEBAPT--TPKKAPKEBAPTTPKGTSTSDKAPAPT---TPKEBAPTTPKE 740
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1209 ---deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1253
QY 741 PAPPTPKKAPPT--TPETPTPTTPSEVSTPTTKKEPTTIHKSDEPTPELSAEPPTKALENS 799
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1254 ptps--deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdepl 1305
QY 800 PKEGVPPTTKTPAPT--KPEMTTAKDKTERDLTTPETTTAANKMKETATTEKTE 857
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1306 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1361
QY 858 SKITATTTQVSTTTQDTTPPKITTLTKTTTAPKVTTKKITTITTEIMNKEBETAKPRDR 917
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1362 deplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1396
QY 918 ATNSKAT-----TPKQKPTKAP--KKPTSTKKPKMKPVKPKKPTTPTRKM----- 962
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1397 sggsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg 1453
QY 963 -----TSMPELNPSTRIAEAMLQTTTRPNQTPNSKLVENPNSDAGAEGETPHM- 1014
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1454 yllgydgsfipernittraaavf-----akll-----gadesyagsgaspsd 1498
QY 1015 LLRPH-----VMPK-----VTPMDYLRPRVNGI----- 1040
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1499 lachmaawaikfatsgqlfkgydgtfkpqnilttraefatvnlhfltkvvgelmskla 1558
QY 1041 ---IINPMLSDETNICNG---KPVDTLTL-----RNGTLVAFRGHYFMMLSPSPSPS 1087
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1559 tldisnpkfd---cvgnhaqefiekltsyglisgyrdgt-----lfpqn 1600
QY 1088 PARBITEWGSPSPIDYVTRFRCNCEGKTFPPFKD--SQYWRFTINDIKD 1132
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1601 ylkrsesv---allnralerplngapklfpdvnesywf-gdimd 1642

```

```

RESULT 11
AA03645
ID AA03645 standard; Protein: 1325 AA.
XX
XX AA03645;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #2377 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX
XX Inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX MO200157270-A2.

```

```

XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX
XX WPI: 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX
XX In a human breast.
XX
XX Claim 27; SEQ ID NO 12385; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX
XX (see A100010-A110067). The present sequence is a peptide encoded by one
XX
XX such probe. The probes are useful for measuring human gene expression in
XX
XX a human breast sample, where the probe hybridises at high stringency to a
XX
XX nucleic acid expressed in the human breast. The probes are useful for
XX
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX
XX diseases of the human breast, particularly those diseases with polygenic
XX
XX aetiology. The diseases include: breast cancer, disorders of development,
XX
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX
XX breast disease and non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX
XX specification, but was obtained in electronic format directly from WIPO
XX
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1325 AA:

```

```

Query Match 11.0%; Score 776.5; DB 22; Length 1325;
Best Local Similarity 25.9%; Pred. No. 5.8e-37;
Matches 304; Conservative 99; Mismatches 492; Indels 281; Gaps 43;

```

```

QY 76 ESFERGECDDQCKKYKCCDYSPCAEVKDNKKNRTKKKPPKPYVDAGSGLDN 135
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 51 elheqkthcrlthskprtdk--ptgnsktldhksctdn-----eapptseensngk 102
QY 136 GDF-----KVTPDSTTQHNKAVSTSPKITTAA--KPINRPSLPRNSDTSK-----E 180
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 103 dpmllngswdsadstlttkhesagkhhlpapksklnctsktsgkvtlrvksdktgrple 162
QY 181 TSLTVKKEETVETKEETTT-----NKOTSDGKEKTTSAKET-----QSIEKT- 223
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 163 kmsnldktsstshkcttsfmgsgnqtkgsktsfpekltasaktyktgtpseesek 222
QY 224 SAKDLAPTISKVLAKPPTKPKETTTKGPALP-----TPKEBAPT----- 260
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 223 dsrttvaaklllttknigtlsaneltqslaeptehygrtanennnpspeptenrer 282
QY 261 TPKEBAPTTPKEPTPTTISA--PTTPKEBAPT-----TTSKAPTTPKEBAPTTPK 309
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 283 tanenttlspaepnterentanentapfapgtlenemtanentlfpaeptehertan 342
QY 310 E---PAPTTPKAPPTTPKEBAPTTPKSAPTTPKKEBAPTTPKAPPTTPKKEBAPTTPKEP 366
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 343 entcpsaeptehertane-----nltpsaeptehygrtlfandktsssestehgr 398
QY 367 TP-----TPKEBAPT--TPKEBAPTTPKEBAPTAPKAPPTTPKKEBAPTTPKEBAPT--PTT 417
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 399 tplanentpsaeptenrerentanentcpspgeptenrettanenttlispeptenret 458

```

PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Query Match

13.98; Score 981; DB 21; Length 763;  
 Best Local Similarity 38.58; Pred. No. 4.4e-49;  
 Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

QY 243 ETTTGGALTPKEPPTTTPKEPASTTPKEPTTTSAPTTKSAPTTKEPAPTTPKE 302  
 DB 126 qtdagntpylirppprlpppcvctcp-epsr-----pppvtl---pqllpt 170  
 QY 303 PAPTTPKEPAPTTPKEPAPTTPKEPAPT--TKSAPTTPKEPAPTTPKAPPTKEPAP 360  
 DB 171 lprtlpctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 226  
 QY 361 TTPKEPAPTTPKEPAPT--KEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPAPT 419  
 DB 227 ntpetpctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 286  
 QY 420 EPSPTTPKEPAPTTPKSAPTTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKE 479  
 DB 287 etspntppctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 340  
 QY 480 PKKAPPTTPKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKAPPTTPKEPAPTTP 539  
 DB 341 ppgpntpppctpppctpppctpppctpppctpppctpppctpppctpppctpppct 400  
 QY 540 APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKE 599  
 DB 401 lprtlpctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 460  
 QY 600 PKKAPPTTPKEPAPTTPKGTAPPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPAPT 659  
 DB 461 pklp-ptlppctpppctpppctpppctpppctpppctpppctpppctpppctpppct 514  
 QY 660 TPKGTAPPTTPKEPAPTTPKEPAPTTPKGTAPPTTPKEPAPTTPKAPPTTPKEPAPT 719  
 DB 515 tppgtpppctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 574  
 QY 720 TTSDKP-----APTTPKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKEPAPTTP 772  
 DB 575 lprtlpctpppctpppctpppctpppctpppctpppctpppctpppctpppctpp 634  
 QY 773 P---TTIHKSPDSTPELSAETPKALNSKEKGVPTTKKPAATKP 816  
 DB 635 pppttlppspptpppctpppctpppctpppctpppctpppctpppctpppctpppct 681

RESULT 10  
 AAM43106  
 ID AAM43106 standard; Protein; 1664 AA.

AC AAM43106;  
 DT 16-OCT-1998 (first entry)  
 DE C. thermocellum OlpB protein.  
 XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;  
 KW cohesin domain; catalytic subunit; scaffold subunit; sdbd; synergistic;  
 KM cellulosome integrating protein; scaffoldin dockerin binding protein.  
 XX Clostridium thermocellum.  
 OS  
 XX  
 FH Key Location/Qualifiers

FT Domain  
 FT /note="cohesin type II domain"  
 FT 28..192  
 FT Domain  
 FT /note="cohesin type II domain"  
 FT 207..363  
 FT Domain  
 FT /note="cohesin type II domain"  
 FT 409..565  
 FT Domain  
 FT /note="cohesin type II domain"  
 FT 607..763  
 FT Domain  
 FT /note="cohesin type II domain"  
 FT  
 PN FR2748479-A1.  
 XX  
 PD 14-NOV-1997.  
 XX  
 PF 10-MAY-1996; 96FR-0005854.  
 XX  
 PR 10-MAY-1996; 96FR-0005854.  
 XX  
 PA (INSP) INST PASTEUR.  
 XX  
 PI Beguin P, Leibovitz E;  
 XX  
 DR WPI; 1998-011569/02.  
 XX  
 DR N-PSDB; AAT86623.  
 XX

Cellulase proteins with cohesin or dockerin type II domains - useful for potentiating the activity of multiprotein enzyme complexes  
 Claim 7; Page 31-39; 60pp; French.

CC Multimeric protein, especially enzymatic, complexes are held together  
 CC by protein-protein interactions between domains designated dockerins  
 CC and cohesins, which are found on the catalytic and scaffold subunits  
 CC respectively. An example of such a complex is the cellulose degrading  
 CC protein complex from Clostridium thermocellum, known as the cellulosome.  
 CC This complex comprises around 15 proteins including endoglucanases,  
 CC cellobiohydrolases, hemicellulases, e.g. xylnases or lichenases, which  
 CC interact with a central "scaffold" protein designated the cellulosome  
 CC integrating protein (Cipa; see AAM43106). The catalytic subunits  
 CC interact with the Cipa subunit via conserved 23 amino acid dockerin  
 CC domains. Cipa has been shown to contain 9 copies of a cohesin domain.  
 CC The invention relates to the isolation of proteins binding to a novel  
 CC dockerin type domain found in the C-terminal portion of Cipa. The new  
 CC I domain found on the catalytic subunits of the cellulosome) The type  
 CC II dockerin domain has some sequence similarity to the type I dockerins  
 CC but is unable to bind type I cohesin domains.  
 CC The sequence presented here is an example of a protein which binds  
 CC the novel type II dockerin domain and is the product of the OlpB gene.  
 CC The protein contains 4 type II cohesin domains in the N-terminal portion  
 CC of which the first domain (amino acid residues 28-192) is thought to  
 CC bind Cipa. The novel type II dockerin and cohesin domains can be used  
 CC in complexes, especially enzyme complexes, to potentiate their catalytic  
 CC actions in a synergistic manner.  
 CC  
 XX  
 XX Sequence 1664 AA;  
 SQ

Query Match 13.48; Score 950; DB 19; Length 1664;  
 Best Local Similarity 31.08; Pred. No. 6.6e-47;  
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 234 VLAQTP-KAETTTGAPLTPKEPTTPKEPASTTPKEPTTTSAPTTKEPAPTTP 292  
 DB 758 vwidgapikaasdepipcdtspdepips-----depips-----depipsdepips 804  
 QY 293 TKSAPTTPKEPAPTTPKEPAPTTPKEPAPT--PTTKEPAPTTPKSAPTTPKEPAPTTPK 348  
 DB 805 epipsetcepipcltspdepipsdepipsdepipsdepipsdepipsdepipsdepipclt 864  
 QY 349 KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPPTTPKEPAPTTPKE 407  
 DB 865 sdepipsdepips--depips--deftp-sdepipsetcepipcltspdepipsdepips 919

PR	14-MAY-1999	9905-0134219
PR	14-MAY-1999	9905-0134221
PR	14-MAY-1999	9905-0134370
PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0134401
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135533
PR	24-MAY-1999	9905-0135529
PR	27-MAY-1999	9905-0136021
PR	28-MAY-1999	9905-0136392
PR	01-JUN-1999	9905-0136782
PR	03-JUN-1999	9905-0137222
PR	04-JUN-1999	9905-0137528
PR	07-JUN-1999	9905-0137502
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	14-JUN-1999	9905-0138847
PR	16-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	17-JUN-1999	9905-0139453
PR	18-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	22-JUN-1999	9905-0139817
PR	23-JUN-1999	9905-0139899
PR	23-JUN-1999	9905-0140353
PR	24-JUN-1999	9905-0140354
PR	28-JUN-1999	9905-0140695
PR	30-JUN-1999	9905-0140823
PR	01-JUL-1999	9905-0140991
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142054
PR	08-JUL-1999	9905-0142390
PR	09-JUL-1999	9905-0142802
PR	12-JUL-1999	9905-0142920
PR	13-JUL-1999	9905-0142977
PR	15-JUL-1999	9905-0143542
PR	15-JUL-1999	9905-0143624
PR	16-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144086
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144864
PR	21-JUL-1999	9905-0144814
PR	21-JUL-1999	9905-0145086
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145085
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145152
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224

PR	26-JUL-1999;	9905-0145576
PR	27-JUL-1999;	9905-0145513
PR	27-JUL-1999;	9905-0145918
PR	27-JUL-1999;	9905-0145918
PR	28-JUL-1999;	9905-0145519
PR	28-JUL-1999;	9905-0145551
PR	02-AUG-1999;	9905-0146386
PR	02-AUG-1999;	9905-0146388
PR	03-AUG-1999;	9905-0146389
PR	03-AUG-1999;	9905-0147308
PR	04-AUG-1999;	9905-0147302
PR	05-AUG-1999;	9905-0147192
PR	05-AUG-1999;	9905-0147260
PR	06-AUG-1999;	9905-0147303
PR	06-AUG-1999;	9905-0147316
PR	09-AUG-1999;	9905-0147493
PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148317
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148341
PR	13-AUG-1999;	9905-0148565
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0149368
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0148426
PR	20-AUG-1999;	9905-0148722
PR	20-AUG-1999;	9905-0148723
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149902
PR	25-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150884
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	27-AUG-1999;	9905-0151080
PR	30-AUG-1999;	9905-0151080
PR	31-AUG-1999;	9905-0151303
PR	01-SEP-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0150363
PR	10-SEP-1999;	9905-0150730
PR	13-SEP-1999;	9905-0153788
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154779
PR	22-SEP-1999;	9905-0155139
PR	23-SEP-1999;	9905-0155486
PR	24-SEP-1999;	9905-0155486
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156596
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157753
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158222
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0159293
PR	13-OCT-1999;	9905-0159294
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159330
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159637
PR	18-OCT-1999;	9905-0159684
PR	18-OCT-1999;	9905-0159684
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160770
PR	21-OCT-1999;	9905-0160814
PR	22-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160988

(e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512  
CC and AA24494 to AA24523 represent nucleotide and amino acid sequences  
CC given in the exemplification of the present invention.

XX Sequence 5179 AA;

Query Match 16.7%; Score 1180; DB 22; Length 5179;  
Best Local Similarity 30.4%; Pred. No. 1, le-59;  
Matches 389; Conservative 68; Mismatches 512; Indels 310; Gaps 48;

QY 4 KTLPIYLLLLSVFIQOVSSQDLSSC--AGRCG-----EGISRDATCNCDYQOHMEC 56  
DB 1169 kdrply-----eedlkkcvadkcyvedthypgagavpleetckscv-c 1213  
QY 57 CDPFKVCYCNELSCGRCEFSFERG-----REGDCDAQCKKYDKC-----PDYES 102  
DB 1214 tnsqyvcvcre---egllnqtdgafcywelcpgnvtckhnicstlrpsltltft 1270  
QY 103 FCAEVKDNKKKRTKKKPPKPVV-----DEAGSGLDNGDEK----- 139  
DB 1271 itlptprtstttttttstestvstfplclclwsdwnedhpsgsdgdrefqgvc 1330  
QY 140 -----VTPPDSTTOH-NKVSTSPK----- 158  
DB 1331 gapedleecrsvkkqphlslehgqgkvqcdvsvgficknedgfgnpgfqlgydklrvncw 1390  
QY 159 -----ITAPKINRPSLPPNSDTSKETSLVNKETVETKETTNTKQTSDEKERTS 213  
DB 1391 pmdcltspsttspsttspsttltlpttspstt-----tltppttspstt 1445  
QY 214 AKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPALETTPKEPT-----PTTPKEPAST 268  
DB 1446 ttp-----lpt-----tpspstlttltpttspsttspsttspsttspstt 1489  
QY 269 TPKEPTPTTIKSAP-TPKPEP-----APTTKSAP-TPKPEPAPTTKPEAPTTKPEP 319  
DB 1490 tttpttpttspsttspsttspsttltlpttspsttlttltpttspsttptltp 1549  
QY 320 -----APTTPKPEAPTTTKSAP--TPKPEAPTPPKPAPTPKPEAPTTKPEPTPT 370  
DB 1550 pttstlttpttspsttspsttlttpttspsttspsttspsttlttpttspstt 1609  
QY 371 PKPEAPTTKPEAPTTKPEP-----APTAKKPAPTPKPEAPTTKPEAPTTKPEPSP 423  
DB 1610 ttppttspsttpttltptstltlpttspsttlttpttspsttspsttspstt 1669  
QY 424 TPKEPAPTTTKSAPTTKPEAPTTTKSAP--TPKPEPPTTKPEAPTTKPEAPTT 480  
DB 1670 ttttpttpttspsttspsttlttpttspsttlttpttspsttlttpttspstt 1728  
QY 481 KKPAPTTKPEAPTTKPEAPTTTKKPAPTAKPEAPTT-----KEPAPTT----- 528  
DB 1729 ttpsttspsttpttlttpttstltlpttspsttlttpttspsttlttpttspstt 1788  
QY 529 -----KKLTP----- 533  
DB 1789 tglwldsgkpnfhkpgdteiligdcvgpwaaniscratmnpdvrgqlgqvtvcvdsvgl 1848  
QY 534 -----TPPKLAPTTPEKPAPTPEELAPTTPEE 562  
DB 1849 lcknedqkpgyvimafclnyelngvcecvqptmttttlenpcttpttlttltvt 1908  
QY 563 PTPPT-TPPEPAPTTPKAAPNTPKPEAPTTKPEP-----APTTPKEPAPT-TPKPEAPTT 615  
DB 1909 ptpctpgtqtpcttpttlttvtcprpctgqtpcttpttlttvtcprpctgqtpctt 1968  
QY 616 PKGTAPTTLKEPAPTTPKKPAKELAPTTTKETSTTSUKPAP-TPPKGTAPTTKPEAP 674  
DB 1969 ptttcttvtcprpctgqtpcttpttlttvtcprpctgqtpcttpttlttvtcprp 2028  
QY 675 TPPEKAPTTPKGTAPTTLKEPAPTTKKKPAKELAPTTTKAPTSTTSUKPAP-TPPKPT 733  
| : | | | | | : | : | : | | | | |

DB 2029 tqtqtp-ttptlttltvtcprpctgqtpcttpttlttltvtcprpctgqtpcttptlt 2087  
QY 734 APTPKPEAPTTPKKPAPTTPTPTPTTSEVSTPTTKPTTIHNSPDESTPELSAEP 793  
DB 2088 tttvtcprpctgqtpctt-ttptlttltvtcprpctgqtpctt--tplttltvtcprp 2143  
QY 794 KALENSPKPEP-----GVPET-TPKPAATPKPEKTTAKDKTTERDLR--TTP--ET 837  
DB 2144 tqtqtpcttpttlttltvtcprpctgqtpcttpttlttltvtcprpctgqtpcttptlt 2203  
QY 838 TPAPKMT--KETATPTEKTESKITATTOVSTTODTTPFKITILKITTLAKVYT 894  
DB 2204 tvtcprpctgqtpcttpttlttltvtcprpctgqtpctt--ltt--lttvtcprp 2258  
QY 895 T-KKITTTTETIMKPEETAKPKDRATNSKATPKPKQKPTKAP-KKPTSTKPKTPRVRK 952  
DB 2259 tqtqtpcttpttlttltvtcprpctgqtpcttpttlttltvtcprpctgqtpcttptlt 2318  
QY 953 PKTTPPKKMTSMPELNP--TSRIAEAMLQTTTP-PNOTPNSKLVEVNPKSEDAAGAE 1008  
DB 2319 tvtcprpctgqtpcttpttlttltvtcprpctgqtpcttpttlttltvtcprpctg 2375  
QY 1009 GEPFHMLRPHVPEVTP 1027  
DB 2376 tqtq--ttptlttltvtp 2392  
| : | | | | | | | | | |

RESULT 9  
AAG38942 ID AAG38942 standard; Protein; 763 AA.  
XX AAG38942;  
AC  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 99US-0134218.



	Query Match	24.2%	Score 1707.5;	DB 22;	Length 472;
	Best Local Similarity	79.9%;	Fred. No. 2e-91;		
	Matches 319;	Conservative	24;	Mismatches 53;	Indels 3; Gaps 3
QY	914 PKDRATNSKATTPPKPQKEFKAPRKPTSTGKKPTMPRRVRKKETPPRPKR-MTSMPELNPT				972
Dd	76 pkyratnsqvtctppkqpcktparkpkstckrpt-privatektltpcpkttcstampcptc				134
QY	973 SRIAAAIQTTRPRQGTNSKLVEYNPKSESDAGAEGETPHMLLRPHVFMPEVTIPDMYL				1032
Dd	135 s-lpeamqlqtctrpnpnselldvnsenedgdaaegekrhlffrpvlrlpvlgteii				193
QY	1033 PRVPNOGIILINMPLSDENPNICNGKRGVDGTLTTRNCTIIVAFRGHYMMMSPEPSPARR				1092
Dd	194 vrgpsqfgslmpmfcdencnlgprvgdlitclngllvalfghytwmltfricpppprr				253
QY	1093 TEWVGISPPIDTVLFRCKCEGKTFFEKDSQYWRFTNDIKDACYPAKPIFGFGLTGQIYA				1152
Dd	254 tevwglspidltvlfrckceegktffifksgywrftndikdagypkliskfgslngkiva				313
QY	1153 ALSTAKYNNMESESYEFKRGSGIOOYIKOEYGVCKPCRRPALANPVYGEMIQVRRRE				1212

Db	314	alslslagyskrpsesvfyfkrkgsvsqgqlytkkqdebtckcttrrpahnsyvygetavrrrrife	373
Qy	1213	RAIGSQTHTIRIQYSPARLAVODKGVLIHNEKVSILMKGLEPNVYTSALSPNIRKPDGY	1272
Db	374	raigsgqvthlrlhytlpvrvpfdkgflhnevkastlwrjlpnvvtssalslpnlkpkdgy	433
Qy	1273	DYAFSPKDDQYINIDVPSRPARVIRTRSGQTSKVMYVNC	1311
Db	434	dyafalskdqyynidvpsrparvtrairltrsrqqlsnwncp	472
	RESULT		
	8		
ID	AAM24516		
XX	AAM24516	standard; Protein: 5179 AA.	
AC	AAM24516;		
XX			
DT	12-OCT-2001	(first entry)	
XX			
DE	C899P	predicted amino acid sequence.	
XX			
KW	Human; immunotherapy; diagnosis; colon cancer; colon tumour;		
KW	immunogenic; gene therapy; vaccine; colonic cancer.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200149716-A2.		
XX			
PD	12-JUL-2001.		
XX			
FE	29-DEC-2000; 2000WO-US35596.		
XX			
PR	30-DEC-1999; 99US-0476296.		
PR	10-JAN-2000; 2000US-0480321.		
PR	15-FEB-2000; 2000US-0504629.		
PR	06-MAR-2000; 2000US-0519444.		
PR	19-MAY-2000; 2000US-0575251.		
PR	29-JUN-2000; 2000US-0609448.		
XX	28-AUG-2000; 2000US-0649811.		
XX			
PA	(CORI-) CORIXA CORP.		
PI	Xu J, Lodes MJ, Secretst H, Benson DR, Meagher MJ, Stolk JA;		
PI	King GE, Wang T, Jiang Y;		
XX			
DR	WPI; 2001-441847/47.		
XX			
PT	Colon tumor associated proteins and nucleic acids useful for the		
PT	prevention, diagnosis and treatment of colonic cancer -		
XX			
PS	Claim 2; Page 446-462; 472pp; English.		
XX			
CC	The present invention describes colon tumour associated proteins (I) and		
CC	the polynucleotides (II) that encode them. (I) have cytostatic activity		
CC	(I) and (II) can be used in gene therapy and vaccine production. (I) and		
CC	(II) may be used in the prevention, diagnosis and treatment of diseases		
CC	associated with inappropriate colon tumour associated protein (TCAP)		
CC	expression, such as colonic cancer. For example, (I) and (II) may be		
CC	used to treat disorders associated with decreased expression by		
CC	rectifying mutations or deletions in a patient's genome that affect the		
CC	activity of TCAPs by expressing inactive proteins or to supplement the		
CC	patients own production of them. Additionally, (II) may be used to		
CC	produce the TCAP protein, by inserting the nucleic acids into a host		
CC	cell, culturing the cell to express the protein. (II) and its		
CC	complementary sequences may also be used as DNA probes in diagnostic		
CC	polymerase chain reaction (PCR) and hybridisation assays to detect and		
CC	quantitate the presence of similar nucleic acids in samples, and		
CC	therefore which patients may be in need of restorative therapy. (I) may		
CC	also be used as antigens in the production of antibodies against TCAPs		
CC	and in assays to identify modulators of TCAP expression and activity.		
CC	Anti-(I) antibodies and antagonists may also be used to down regulate		
CC	TCAP expression and activity. The anti-(I) antibodies may also be used		
CC	as diagnostic agents for detecting the presence of TCAPs in samples		



[illegible]

(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
Jay GD;  
WPI; 2001-024673/03.  
Novel triboneectin polypeptide useful as lubricant for treating  
osteoarthritis, comprises O-linked lubricating moiety -  
Disclosure; Fig 1; 47pp; English.

The invention relates to a human tribonecstin which is a product of alternative splicing of the human MSF (megakaryocyte stimulating factor) gene. The tribonecstin has at least one O-linked oligosaccharide lubricating moiety and has a polypeptide sequence comprising 1-76 (repeats of a motif having at least 50% identity to the sequence KEAPTPT (AA829774)). The invention also relates to a nucleic acid encoding a human MSF-derived tribonecstin; a biocompatible composition comprising a human tribonecstin for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonecstin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonecstin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is epichordal tissue. DNA encoding a tribonecstin may be used in gene therapy. The present sequence represents a substantial portion of a human MSF-derived tribonecstin.

SQ Sequence 902 AA;

Query Match	49.3%	Score 3484;	DB 22;	Length 902;
Best Local Similarity	74.6%	Pred. No. 4.9e-194;		
Matches 794;	Conservative 28;	Mismatches 80;	Indels 162;	Gaps 66

QY	107	VKKKKNNNTKKKPPKPPVYDGAASSGLDNGDFKVTTPDTSTTOHNVKSPTKTTAAPI	166
Dp	1	VKKKKNNNTKKKPPKPPVYDGAASSGLDNGDFKVTTPDTSTTOHNVKSPTKTTAAPI	60
QY	167	PPSLPPNSDSKESLTVNKEETVEERKEHTTTKKONSTOGKEKTSKAKTOSIEKKS	228
Dp	61	PPSLPPNSDSKESLTVNKEETVEERKEHTTTKKONSTOGKEKTSKAKTOSIEKKS	120
QY	227	DLAPTSVLAKEPTPKAETTTKGALTTTPKKEPTTTKPEASATTPKKEPTTTIKSAPTT	286
Dp	121	DLAPTSVLAKEPTPKAETTTKGALTTTPKKEPTTTKPEASATTPKKEPTTTIKSAPTT	172
QY	287	EPAPTTTKSAPTTPKKEAPPTTTKEAPATTTKEAPATTTKKAAPTTTKSAPTTKPEAPTT	346
Dp	173	EPAPTTTKSAPTTPKKEAPPTTTKEAPATTTKEAPATTTKKAAPTTTKSAPTTKPEAPTT	228
QY	347	PKKAPPTTPKPEAPPTTKKEPTPTTPPKPEAPPTTKPEAPPTTKPEAPPAKKAPPTPKPEA	406
Dp	229	PKKAPPTTPKPEAPPTTKKEPTPTTPPKPEAPPTTKPEAPPTTKPEAPPAKKAPPTPKPEA	281
QY	407	PTTPKKEAPPTTKKEPSPTTTKPEAPATTTKSAPTTTKPEAPPTTKKSAPTTKEPSPTTTKE	466
Dp	282	PTTPKKEAPPTTKKEPSPTTTKPEAPATTTKSAPTTTKPEAPPTTKKSAPTTKEPSPTTTKE	335
QY	467	PAPPTPKPEAPPTPKKAPPTPKPEAPPTTKPEAPATTTKKAPATTAKEAPATTPKKAAP	526
Dp	336	PAPPTPKPEAPPTPKKAPPTPKPEAPPTTKPEAPATTTKKAPATTAKEAPATTPKKAAP	388
QY	527	TPKKLPTTPTEKLAPTTPKEKAPATTPKEBELAPTTPEEPTPTPKPEAPPTPKKAAPPTPK	586
Dp	389	TPKKLPTTPTEKLAPTTPKEKAPATTPKEBELAPTTPEEPTPTPKPEAPPTPKKAAPPTPK	440

Accession	Species	Gene	Protein	Length
Dp	Drosophila	ap1	ap1	780
Oy	Oryza	ap1	ap1	747
Db	Drosophila	ap1	ap1	840
Oy	Oryza	ap1	ap1	807
Db	Drosophila	ap1	ap1	900
Oy	Oryza	ap1	ap1	867
Db	Drosophila	ap1	ap1	960
Oy	Oryza	ap1	ap1	927
Db	Drosophila	ap1	ap1	1020
Oy	Oryza	ap1	ap1	987
Db	Drosophila	ap1	ap1	1080
Oy	Oryza	ap1	ap1	1047
Db	Drosophila	ap1	ap1	1140
Oy	Oryza	ap1	ap1	1107
Db	Drosophila	ap1	ap1	1200
Oy	Oryza	ap1	ap1	1167
Db	Drosophila	ap1	ap1	1260
Oy	Oryza	ap1	ap1	1227
Db	Drosophila	ap1	ap1	1320
Oy	Oryza	ap1	ap1	1287
Db	Drosophila	ap1	ap1	1380
Oy	Oryza	ap1	ap1	1311
Db	Drosophila	ap1	ap1	1404
RESULT 4				
AAM24322				
ID	AAM24322	standard; Protein: 1299 AA.		
XX	AC			
XX	AA			
XX	DT	12-OCT-2001 (first entry)		
XX	DE			
XX	XX			
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;			
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;			
KW	diagnostics; forensic test; gene mapping; genetic disorder;			
KW	biodiversity; gene therapy; nutrition.			
XX	OS	Homo sapiens.		
XX	PN	WO20015477-A2.		
XX	PD	02-AUG-2001.		
XX	PF	25-JAN-2001; 2001WO-USO2687.		
XX	PR	25-JAN-2000; 2000US-0491404.		
XX	PR	17-JUL-2000; 2000US-0617746.		

03-AUG-2000; 2000US-0631451.  
15-SEP-2000; 2000US-0663870.  
(HYSE-) HYSEQ INC.  
Tang YH, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
Cao Y, Dirmancak RA, Zhang J, Werhman T;  
WPI: 2001-476164/51.  
N-PSDB: AAH98981.  
Isolated polypeptide for treatment of diseases, diagnostics, raising  
antibodies and research use -  
Claim 20; Page 1198-1201; 1275pp; English.  
The present invention provides the protein and coding sequences of novel  
proteins from a variety of organisms, including human, dog, cat, horse,  
cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
urchin and tomato. These were derived from expressed sequence tags (ESTs)  
from the organism of interest. They can be used in diagnostics,  
forensics, gene mapping, identification of mutations, to assess  
biodiversity and for nutritional purposes. The present sequence is a  
protein of the invention.

	Query Match	91.0%;	Score 6431.5;	DB 22;	Length 1299;	
	Best Local Similarity	92.8%;	Pred. No. 0;			
	Matches 1202;	Conservative	0;	Mismatches	0;	Indels 93; Gaps 17;
QY	1	MAKMTLPITYLLLSFYVLIQOVSSODLSSACAGRCGEYSRDATMCNDYNCOHMECCPDF	60			
Db	1	makmtlpilyllllslvfviqyvsqgdlssacagrcgsqyrdatcnodcnyymecppif	60			
QY	61	KRVCTAELSCGRGCFESFERGERECDDAOCKKYDKCCPDYESCAE-----	106			
Db	61	krvctaelsckgrcfesfergeredcdackkydkccpdyesfcaevhnptpsaskap	120			
QY	107	-----	106			
Db	121	pqsasqlsttkrkspkpnkkklkvieseeitehsnsengasssssssstliw	180			
QY	107	-----VKDNKKNRTKKKPPRPKPVVDEAGSGLDNGFKYTTVDYST	147			
Db	181	kikskksnaanelqkklikvkdnkntkkkpkrpvdeaaqsgldngdfxvtltpdst	240			
QY	148	TQHNVSTSKITTAIPINRPSLPRNSDVSKESTSIVLNKETTVETIKETTYTKQFSTG	207			
Db	241	tqhmvstspkltakipinprslprnsdsketslvnketlevkelttnkgfstg	300			
QY	208	KEKTSAKENOSIEKTSAKDIAPT SKVLAPTPKAETTTKGPAITTPKEPTPTTPEKPS	267			
Db	301	kekttsakeqselektasakladipskylakrpbkaettkgpalttpeptltpkepas	360			
QY	268	TTTTKEPPTTIKSAPTPKEBPATTTKSAATTTKEBAPTTTKBPAPTTKEPATTTKEP	327			
Db	361	tttpeptlttksaptpckepaplttksaptpckepaplttkcepaplttkcpaplttkcp	420			
QY	328	APTTSASPTTPKEBAPTTPKRPAPTTTPKRPAPTTPEBPTPTTPKEBAPTTKEBAPTTPK	387			
Db	421	apttksaptpckepaptpckpapttkpapttkcpaplttkcpaplttkcpaplttkcpaplttk	480			
QY	388	EPAFTAKKRAPPTPKBPAPTTPKEBAPTTTKESPPTTPKEBAPTTTKSAPTTTKBPAPT	447			
Db	481	epaftakkrpapttkbpaptpckpapttkbpspfttkbpaplttksaplttkcaplttkcap	540			
QY	448	TTTGAPTTPKPSPTTTKEBAPTTPKBPAPTTTPKBPAPTTTPKBPAPTTPKBPAPTTTKP	507			
Db	541	tttgaptpkpspttkcpaptpckpapttkbpaptpckpapttkbpaptpckpapttkcp	600			
QY	508	APTAPKEBAPTTPKEIATAITPKILPTTPEKLAFTTBKRAPTTPEBLAFTTPEBPTPTT	567			

QY 928 PQRPAKPKPTSTKPKPTMPVRVRRKPTPTPRKMTSTMBELNPTSRIAEAMLOTTTBN 987  
DB 1021 PQRPKAPKPKPTSTKPKPTMPVRVRRKPTPTPRKMTSTMBELNPTSRIAEAMLOTTTBN 1080  
QY 988 QTPNSLVEVNPSEDAAGAGETPMMLRPHVMEVPTPMDYLRPNOCIIINPMLS 1047  
DB 1081 QTPNSLVEVNPSEDAAGAGETPMMLRPHVMEVPTPMDYLRPNOCIIINPMLS 1140  
QY 1048 DETNINCGRPVDDLTTLRNGTTLVAFRGHYFMMLSPSPSPARTRTEVWGICSPIDTPT 1107  
DB 1141 DETNINCGRPVDDLTTLRNGTTLVAFRGHYFMMLSPSPSPARTRTEVWGICSPIDTPT 1200  
QY 1108 RCMCEKTEFFKDSQYWRNTNDIKDAGYRPIFKGFGSLTGOIYVALSTAKYKMPESY 1167  
DB 1201 RCMCEKTEFFKDSQYWRNTNDIKDAGYRPIFKGFGSLTGOIYVALSTAKYKMPESY 1260  
QY 1168 FFKRGSIQYTYKQEPVQKCPGRPALNYPYGEVTOVRRRRFRERAGPSTHTIRIQ 1227  
DB 1261 FFKRGSIQYTYKQEPVQKCPGRPALNYPYGEVTOVRRRRFRERAGPSTHTIRIQ 1320  
QY 1228 SPARLAYODKGVLAHNEVKYSILMRGLPNVVTSAISLPNIRKPDGYAFSKDYNYNDY 1287  
DB 1321 SPARLAYODKGVLAHNEVKYSILMRGLPNVVTSAISLPNIRKPDGYAFSKDYNYNDY 1380  
QY 1288 PSRTARATTRSGOTLSKRWYNCP 1311  
DB 1381 PSRTARATTRSGOTLSKRWYNCP 1404

RESULT 3  
AAB29773  
ID AAB29773 standard; Protein: 1404 AA.  
XX  
AC AAB29773;  
XX  
DT 28-FEB-2001 (first entry)  
XX  
DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.  
XX  
KW Human MSF, megakaryocyte stimulating factor; tribonectin;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KW friction coefficient reduction; gene therapy; antiarthritic;  
KW osteopathic.  
XX  
OS Homo sapiens.  
XX  
PN WO200064930-A2.  
XX  
PD 02-NOV-2000.  
XX  
PE 24-APR-2000: 2000MO-US10953.  
XX  
PR 23-APR-1999: 99US-0298970.  
XX  
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
XX  
PI Jay GD;  
XX  
DR WPI: 2001-024673/03.  
XX  
N-PSDB: AAC81498.  
PT Novel tribonectin polypeptide useful as lubricant for treating  
PT osteoarthritis, comprises O-linked lubricating moiety  
XX  
PS Claim 3; Page 7; 47pp; English.  
XX  
CC The invention relates to a human tribonectin which is a product of  
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonectin has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76  
CC repeats of a motif having at least 50% identity to the sequence KEPAPTT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a

CC human MSF-derived tribonectin; a biocompatible composition comprising a  
CC human tribonectin for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC tissues of a mammal, where the injury is caused by a surgical insertion  
CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonectin may be used in gene therapy. The present sequence represents  
CC human MSF.  
XX  
SQ Sequence 1404 AA:

Query Match 99.28; Score 7007.5; DB 22; Length 1404;  
Best Local Similarly 93.4%; Pred. No. 0;  
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQOVSSODLSGAGRGEGYSRDATNCNCOHTWECCPDF 60  
DB 1 MAWKTLPIYLLLSVFIQOVSSODLSGAGRGEGYSRDATNCNCOHTWECCPDF 60  
QY 61 KRVCTAELSCGRCFESFERGECDDAQCCKKYDKCCPDYESFCAE----- 106  
DB 61 KRVCTAELSCGRCFESFERGECDDAQCCKKYDKCCPDYESFCAE----- 106  
QY 107 ----- 106  
DB 107 ----- 106  
QY 121 PPSGASGLIKSTKSPKPKKKKKVLESSEELTEHSVSENSQESSSSSSSSSLW 180  
DB 121 PPSGASGLIKSTKSPKPKKKKKVLESSEELTEHSVSENSQESSSSSSSSSLW 180  
QY 107 -----VKDNKKNRTKKKPKPKPVNDVDEAGSGDNGFKYTTEDTS 147  
DB 181 KIKSKNSAANRELQKLLKVKDKNKTKKKPKPKPVNDVDEAGSGDNGFKYTTEDTS 240  
QY 148 TOHNKYSTSKITPAKINDPSLPPNSDTSKETSILVNKETTVERKETTAKOTSTG 207  
DB 241 TQHNKYSTSKITPAKINDPSLPPNSDTSKETSILVNKETTVERKETTAKOTSTG 300  
QY 208 KEKTSAKETQSTIEKTSAKDLAPTSTVLAKPPTKAETTTGCPALTTPKEPTTKEPAS 267  
DB 301 KEKTSAKETQSTIEKTSAKDLAPTSTVLAKPPTKAETTTGCPALTTPKEPTTKEPAS 360  
QY 268 TTPKEPTTTPKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTK 327  
DB 361 TTPKEPTTTPKSAPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTK 420  
QY 328 APTTTSAPTTPEKPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTPPK 387  
DB 421 APTTTSAPTTPEKPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTPPK 480  
QY 388 EPAPTAAPKAPATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTP 447  
DB 481 EPAPTAAPKAPATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTP 540  
QY 448 TTKSAPTTPEKSPPTTKKAPATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAP 507  
DB 541 TTKSAPTTPEKSPPTTKKAPATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAP 600  
QY 508 APTAPKEPAATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTP 567  
DB 601 APTAPKEPAATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAPPTTPKEPAATTP 660  
QY 568 PEEPAATTPKAAPNTPEKPAATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAP 627  
DB 661 PEEPAATTPKAAPNTPEKPAATTPKEPAATTPPKKAPPTTPKEPAATTPPKKAP 720  
QY 628 APTTPKAPKAPKAPATTTTKEPTSTSDKAPATTPPGTAPTTPEKPAATTPPKG 687

QY 1168 FFKRGSSIOQYIYKQEPVOKCPGRPALNYPVYGEMTOVRRRRRFEERAIGPSQTHIRIOY 1227  
 Db 1261 ffrfgrsgisqgylykqepvokcpgrpalnypvygemtgyrrrrferraigpsqthiriy 1320  
 QY 1228 SPARLAYQDKGVJHNEVKVSIIMRGLPNVYTSALSPNTRKPGCYGYVAFSKOQYNNIDY 1287  
 Db 1321 sparlavgdkgyvhnevksilwrglpsvvtalslpsntrkpgdygyalskqyynidy 1380  
 QY 1288 PSRTARAITRRSGOTLSKVVYNCP 1311  
 Db 1381 psrtaraitrrsgotlskvvynncp 1404  
 RESULT 2  
 AAB60568  
 ID AAB60568 standard; protein; 1404 AA.  
 AC AAB60568;  
 XX  
 DT 27-APR-2001 (first entry)  
 DE Human megakaryocyte stimulating factor (MSF, CACP).  
 XX  
 KM Human: CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;  
 KM MSF; megakaryocyte stimulating factor; synovial lubricant;  
 KM Chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KM antiarthritic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W020107068-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 21-JUL-2000; 2000MO-US20002.  
 XX  
 PR 23-JUL-1999; 99US-0145328.  
 PR 19-JUL-2000; 2000US-0145328.  
 XX  
 PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX  
 PI Warman ML;  
 XX  
 DR WPI: 2001-182721/18.  
 XX  
 PT New composition comprising the campodactyly-arthropathy-coxa  
 PT vara-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints  
 XX  
 PS Example 1; Page -: 34pp; English.  
 CC The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactyly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anaesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactyly-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a Genbank accession number was given. This  
 CC sequence was therefore obtained from Genbank (U70316).  
 XX  
 SQ Sequence 1404 AA;

Query Match 99.2%; Score 7007.5; DB 22; Length 1404;  
 Best Local Similarity 93.4%; Pred. No. 0;  
 Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1.  
 QY 1 MAMKTLPIYLLLLSYFVLIQOVSSODLSSCAGRCGEYSRDATCMCDYINCHMYECCDF 60  
 Db 1 mawktlpiyl11111svfviqyvsqdlsscagrcgeysrdatcmcdyncchymeccpdf 60  
 QY 61 KRYCTAELSCGRCFESFGRRECCDACCCKYDCCPPYEFSCAE----- 106  
 Db 61 kryctaelsckgrcfesfgrreccdaqckkydcccyyefscavmnpyspssckap 120  
 QY 107----- 106  
 Db 121 ppsgaqltksttkrpkppnkkkklvleseeleehsvsengeassssssssstliw 180  
 QY 107-----VKDNKKNRYYKKKPPKPPVYDEAGSGLDNDFKVTTPDTS 147  
 Db 181 klksksnaanelqkklkvkdnknrtkkkppkppvdeagsgldngdfkvtlpdst 240  
 QY 148 TQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSILVNNKETTERTTTNNKQTSIDG 207  
 Db 241 tqhnkvstspkiltakpimprslppnsdtsketsilvnnkettvektettunkqtsidg 300  
 QY 208 KERTTSAKEROSIEKTSANDLAPTSKVLAKPPKAEFTTGKPALETTPKKEPTTPKEPAS 267  
 Db 301 kektsakerqsietsakndlaptskvlakppkaelttgkpalettppkepttpkpas 360  
 QY 268 TTPKEPTPTTIKSAPTTPKKEPAATTTKSAPTTPKKEPAATTPKKEPAATTTKKEP 327  
 Db 361 ttpkeptpttiksaptpkkepalttksaptpkkepalttksaptpkkepalttkkep 420  
 QY 328 APPTTKSAPTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPK 387  
 Db 421 apttksaptpkkepalttksaptpkkepalttksaptpkkepalttkkepalttkp 480  
 QY 388 EPAPTAERKAPPTPKKEPAATTPKKEPASPTTPKKEPAATTTKSAPTTKKEPAPT 447  
 Db 481 epaptakrkaptpkkepalttkkepalttkkepalttkkepalttksaptpkkepalt 540  
 QY 448 TTKSAPTPKKEPSPTTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEP 507  
 Db 541 ttksaptpkkepalttkkepalttkkepalttkkepalttkkepalttkkepalttkp 600  
 QY 508 APAPKKEPAPTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 567  
 Db 601 apapkepapttkkepalttkkepalttkkepalttkkepalttkkepalttkp 660  
 QY 568 PEEPAATTPKKAANTREKREPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEP 627  
 Db 661 peepapttkkaantrekrepalttkkepalttkkepalttkkepalttkpalttkp 720  
 QY 628 APPTPKKAPKELAPPTTKKEPTSTSDKAPATTPKGAATTPKKEPAATTPKKEP 687  
 Db 721 aptpkkapekelpapttkkeptsstsdkapattpkgaattpkkepalttkp 780  
 QY 688 TAPPTLKEPADTPPKKAPKELAPPTTKGPTSTSDKAPATTPKKEPAATTPKKEP 747  
 Db 781 taptlkepadtpkkapekelpapttkgptstsdkapattpkkepalttkp 840  
 QY 748 KPAPPTTPPTPTSEVSTPTTKKEPTTIHKSPESTPELSAEPPKALSPSPGCVPT 807  
 Db 841 kpapttptptptsevspttkkepttihspspeelsaeppkalspspgcvpt 900  
 QY 808 TTPPATKPEMTTAKDPTTERDLKTPPETTTAAPKMTKETATTEKTTATTTGY 867  
 Db 901 tkpatkpemttakdptterdlktppetttaapkmktetatekttatlatgy 960  
 QY 868 TSTTTQDTPPKITTLTKTTTAPKVTYTKKITTTEINMKKEEPAKPKDRATNSATTPK 927  
 Db 961 tstttqdtppkittltktttlapkvttytkkittteimkpeeakpkdratnsattpk 1020



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:34 ; Search time 138.84 Seconds  
(without alignments)  
699.439 Million cell updates/sec

Title: AAB  
Perfect score: 7064  
Sequence: 1 MAWKTPILYLLLSVFVQ.....AAITRSGQLSKWYNCP 1311

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101:\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7007.5	99.2	1404	13 AAR26049	MSF precursor. SY
2	7007.5	99.2	1404	22 AAB60568	Human megakaryocyt
3	7007.5	99.2	1404	22 AAB29773	Human megakaryocyt
4	6431.5	91.0	1299	22 AAM24322	Human EST encoded
5	3484	49.3	902	22 AAB29778	Human MSF-derived
6	1934.5	27.4	452	16 AAR80041	Human megakaryocyt
7	1707.5	24.2	472	22 AAB60569	Bovine MSF ortholo
8	1180	16.7	5179	22 AAM24516	C899P predicted am
9	950	13.9	763	21 AAG38942	Arabidopsis thalia
10	950	13.4	1664	19 AAM43106	C. thermocellum OI
11	776.5	11.0	1325	22 AAM03645	Peptide #2327 enco

12	715	10.1	763	18 AAM31852	Mycobacterium tube
13	627	8.9	4412	21 AAY53666	Sequence g1/101742
14	612	8.7	572	18 AAM31855	Mycobacterium tube
15	547.5	7.8	844	7 AAB60570	Sequence of the Pa
16	542	7.7	807	21 AAY54467	Aminl acid sequenc
17	529	7.5	182	12 AAR10872	Protein encoded by
18	520.5	7.4	788	21 AAY54466	Amino acid sequenc
19	508	7.2	1837	21 AAB11726	Cryptosporidium pa
20	506.5	7.2	744	9 AAB2975	Bioadhesive precur
21	498	7.0	111	13 AAR26050	MSF-K130. Synthet
22	496.5	7.0	2971	21 AAB41231	Human ORFX ORP95
23	489.5	6.9	2972	22 AAB50363	Human SRCAP. Homo
24	489.5	6.9	3118	22 AAB50362	Portion of Cryptos
25	489	6.9	1721	21 AAB11727	P. Yoelli SSP2 ant
26	488.5	6.9	826	13 AAR6042	Peptide #2892 enco
27	488	6.9	617	22 AAM16458	Peptide #2869 enco
28	488	6.9	617	22 AAM04187	Human MUC11 polype
29	488	6.9	957	21 AAY59288	C900P predicted am
30	488	6.9	957	22 AAM48299	Cryptosporidium pa
31	488	6.9	1721	19 AAM48299	Human protein sequ
32	485	6.9	1127	22 AAB95541	Bioadhesive precur
33	476.5	6.7	652	9 AAB2974	Peptide #1317 enco
34	467.5	6.6	511	22 AAM14883	Peptide #1349 enco
35	467.5	6.6	511	22 AAM27312	Peptide #1289 enco
36	467.5	6.6	511	22 AAM02607	C. albicans Rot1 p
37	456.5	6.5	750	20 AAY05477	PRP 378. Trilicun
38	450.5	6.4	378	12 AAR14160	PRP encoded by clo
39	446.5	6.3	378	12 AAR14162	Caenorhabditis ele
40	443	6.3	2870	21 AAY95559	Caenorhabditis ele
41	443	6.3	3178	21 AAY95556	Japanese sea musse
42	442.5	6.3	751	16 AAR80839	Human OTCG27 gene
43	437.5	6.2	2819	22 AAB35408	KTFP-DEVD-MAP4-EBF
44	429	6.1	1610	21 AAB22870	Bifunctional caspa
45	429	6.1	1610	21 AAY79648	

## ALIGNMENTS

RESULT	ID	AA	Location/Qualifiers
1	AAR26049	standard; Protein; 1404 AA.	
AC	AAR26049;		
XX	02-FEB-1993 (first entry)		
DT	XX	MSF precursor.	
XX	XX	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;	
XX	KW	stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	OS	Synthetic.	
XX	XX	Key	
FT	FT	Region	1..26
FT	FT	Region	/label= Exon_I
FT	FT	Region	26..67
FT	FT	Region	/label= Exon_II
FT	FT	Region	67..107
FT	FT	Region	/label= Exon_III
FT	FT	Region	107..157
FT	FT	Region	/label= Exon_IV
FT	FT	Region	157..200
FT	FT	Region	/label= Exon_V
FT	FT	Region	200..1141
FT	FT	Region	/label= Exon_VI
FT	FT	Region	1411..1166
FT	FT	Region	/label= Exon_VII
FT	FT	Region	1166..1212
FT	FT	Region	/label= Exon_VIII
FT	FT	Region	1213..1266
FT	FT	Region	/label= Exon_IX

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the Pa  
Aminl acid sequenc  
Protein encoded by  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precur  
MSF-K130. Synthet  
Human ORFX ORP95  
Human SRCAP. Homo  
Human SRCAP. Homo  
Portion of Cryptos  
P. Yoelli SSP2 ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 polype  
C900P predicted am  
Cryptosporidium pa  
Human protein sequ  
Bioadhesive precur  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
C. albicans Rot1 p  
PRP 378. Trilicun  
PRP encoded by clo  
Caenorhabditis ele  
Caenorhabditis ele  
Japanese sea musse  
Human OTCG27 gene  
KTFP-DEVD-MAP4-EBF  
Bifunctional caspa



p. 1-3

















PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132487.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134270.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134368.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142970.  
PR 12-JUL-1999; 99US-0142970.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154019.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0156359.  
PR 29-SEP-1999; 99US-0156458.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512  
 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences  
 CC given in the exemplification of the present invention.

XX Sequence 5179 AA:

Query Match 16.1%; Score 1175; DB 22; Length 5179;  
 Best Local Similarity 29.3%; Pred. No. 7.1e-56;  
 Matches 398; Conservative 81; Mismatches 519; Indels 360; Gaps 52;

```

QY 36 SFEGRGRCDC-----DAQCKKYDKC-----CPDSESCAEVHNPTSPSSKKAP 79
DB 1081 scdtggcdectcsavayaqecckegacvfrtpdlcp---lfc-dyynphceewhyep 1136
QY 80 PPSGASQTIKSTT-----KRSPP--PPNKKTKKVIIESE-----ITEEH- 117
DB 1137 cgnstfctortlinghanisvysylegcyrcpkdrplyeodlkvcvackcgcyvedlhy 1196
QY 118 ---SVSENOESSSSSSSSSSSTT-----WKIKSSKNSAANRELQK 155
DB 1197 ppgasvpleetckscvctnssqvvcreegklnqtdgafcyweI---cgnpgtvek 1251
QY 156 -----KLKVKDNKKNKTKKPPKPPV----- 178
DB 1252 hfnlcstlrpsltlfttlrlptlpsfttlcttptststvlstplcklclawdwlne 1311
QY 179 DEAGSGLDNGDFK-----VTPPTSTTQH-NKYSTSK----- 210
DB 1312 dhpsgsddgdrpfdyvcgapediecrsvkdpnlslqhqkqvcduvsygfickneqf 1371
QY 211 -----ITTAKPINRPSLPPNSDTSEKSTSLVNNKETVETK 246
DB 1372 gngpfqldydklyncwcmndkcltspptlpsppptlcllptlpsptt----- 1427
QY 247 ETTTNNKQSTDGKEKTSANETOSIEKTSAKDLAPTSKVLAKPPKAEITTKGHALTP 306
DB 1428 -tttppptlpsppitltp-----lpt-----tspstltpptltp 1470
QY 307 KEPT-----PTPKPEAPSTPKEPPTTIKSAP--TPPKP-----APTTSKAP--TTP 352
DB 1471 spptlpsppitlpsppitltpptlpsppmtcltpastltpptlpsptt 1530
QY 353 KEPAFTTKPEAPTPKEP-----APTTEKAPATTTKSAP--TPKEPAFTTKKPA 403
DB 1531 ttpptlpsppitlpsptltpstltpptlpsppptltpptlpsptlpspt 1590
QY 404 PTPKPEAPTPKEPTTPPKPEAPTPPKPEAPTPPKP-----APAPKPEAPTPKE 456
DB 1591 itcttppptlpsppitltpptlpsppitltpstltpptlpsppitltp 1650
QY 457 PAPTTPKEAPTPTKESPSTPKPEAPTPTKSAPTTPKPEAPTTTKSAP--TPKEPSP 513
DB 1651 ppptlpsppitlpsppitltpptlpsppitlpsppitltpmtlpsptlpspt 1709
QY 514 TTTKEAPATPKPEAPTPTKKPAFTTPKPEAPTPPKPEAPTTTKKPAAPKPEAPTPP- 572
DB 1710 ttttptlpsppitlpsptltpptlpsptltpmtltpptlpsptltp 1769
QY 573 ---KETATTP-----KLTLP 585
DB 1770 fspfstltpcltpvcplcnwtglwldsgkpnfhkpggdteligdvcgpygaaanlsrctmp 1829
QY 586 -----TTPKELAPTT 595
DB 1830 dvpigqlgtvcdvsvglcknedkpgsvlpmatclngelavngcccevtqtlmtlct 1889
QY 596 PEKPAFTTPBEIAPTTPEEPTTP--TPEEPAFTTPKAAAPMTPEKPAFTTPKPE-----AP 649
  
```

```

DB 1890 tempptltpitltpitltpvtpvtpvtpvtpitltpitltpitltpvtpvtpvtpvtpitltp 1949
QY 650 TTPKPEAPTP--TPKEAPTPTPKSTATTTKEAPATTPPKKPAKELAPTTTKEPTSTSDKP 708
DB 1950 tttvtpvtpvtpvtpvtpitltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2009
QY 709 AP--TPPKGAFTTPKPEAPATTPKPEAPATTPKSTATTTKEAPATTPPKKPAKELAPTTTK 767
DB 2010 tptltpitltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2068
QY 768 GPTSTTSDKPAP--TPPKPEAPTPTPKPEAPTPPKKAPATTPPEPPTPSEVSPPTTKPEPT 826
DB 2069 tptvtpvtpvtpvtpvtpitltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2127
QY 827 TTHKSPDESPTPELSAPPTKALENSPKPE-----GVPT--TKTPATKPEMTTAK 875
DB 2128 t---tptltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2184
QY 876 DKYTEERDLR---TTP--EPTTAAPKMT---KETATTEKTESKITATTVOYTSITTOOT 927
DB 2185 tptvtpvtpvtpvtpvtpitltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2243
QY 928 TTPKITTTLKTITTLAPKVTIT--KTTITTEIMNKPETAKPKDRATNSKATTPKPKOKPTKA 986
DB 2244 tp---itl---tltvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2299
QY 987 P-KKPTSTKPKMTMPKRVKPKTTPPKMTSTMPELNP--TSRIAEMLQTTTR--PNOT 1041
DB 2300 tptvtpvtpvtpvtpvtpitltpitltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2359
QY 1042 PMSKLVENVNPKSEDDAGAEETPHMLIRPHVPEVTP 1079
DB 2360 pttltpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtpvtp 2392

RESULT 9
AAG38942 standard; Protein; 763 AA.
ID AAG38942
XX
AC AAG38942;
XX
DT 18-OCT-2000 (first entry)
DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990US-0121825.
XX 05-MAR-1999; 990US-0123180.
XX 09-MAR-1999; 990US-0123548.
XX 23-MAR-1999; 990US-0125788.
XX 25-MAR-1999; 990US-0126264.
XX 29-MAR-1999; 990US-0126785.
XX 01-APR-1999; 990US-0127462.
XX 06-APR-1999; 990US-0128234.
XX 08-APR-1999; 990US-0128714.
XX 16-APR-1999; 990US-0129845.
XX 19-APR-1999; 990US-0130077.
XX 21-APR-1999; 990US-0130449.
XX 23-APR-1999; 990US-0130510.
XX 28-APR-1999; 990US-0130891.
XX 30-APR-1999; 990US-0131449.
XX 30-APR-1999; 990US-0132048.
  
```

XX AAB60569;  
AC  
XX  
DT 27-APR-2001 (first entry)  
XX  
DE Bovine MSF orthologue, superficial zone protein (SZP).  
XX  
KW Bovine; CACP protein; camptodactylly-arthropathy-coxa vara-pericarditis;  
XX superficial zone protein; SZP; MSF orthologue; synovial lubricant;  
OS osteoarthritis; joint lubrication; osteopathic; antiarthritic.  
XX Bos taurus.  
XX WO200107068-A1.  
PN  
PD 01-FEB-2001.  
PP 21-JUL-2000; 2000WO-US20002.  
PR 23-JUL-1999; 99US-0145328.  
PR 19-JUL-2000; 2000US-0145328.  
PA (UYCA-) UNIV CASE WESTERN RESERVE.  
PI Warman ML;  
DR WPI; 2001-182721/18.  
XX  
XX New composition comprising the camptodactylly-arthropathy-coxa  
PT vara-pericarditis protein in combination with an anesthetic, useful for  
PT treating osteoarthritis, or as lubricants of tissue and joints -  
XX Example 1; Fig 4; 34pp; English.

PS The invention relates to a method of treating osteoarthritis via the  
XX administration of a composition comprising the camptodactylly-arthropathy-  
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
CC The composition may further comprise a local anaesthetic. The composition  
CC of the invention may be administered via intra-articular or intravenous  
CC injection. The human CACP protein is identified in the invention as  
CC being megakaryocyte stimulating factor (MSF). The gene encoding  
CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
CC this gene are responsible for the heritable disorder camptodactylly-  
CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
CC acts as a synovium lubricant, and can be used to lubricate tissue and  
CC joints in the treatment of osteoarthritis. The composition may be  
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
CC loss of range of movement or joint damage). The present sequence  
CC represents the bovine orthologue of human MSF, superficial zone  
CC protein (SZP).

Sequence 472 AA;  
XX  
XX

Query Match	23.5%	Score 1707.5	DB 22	Length 472
Best Local Similarity	79.9%	Pred. No. 3,7e-86		
Matches 319	Conservative 24	Mismatches 53	Indels 3	Gaps 3
QY 966	PKDRATNSKATTPKFOKPTKABKKPTSTKKRPTMRVRKPKTTPPRK-MTSMPELNP	1024		
Db				
Db	pkgrtngsvttrpkpqrkpktrkpktskkrpt-privrpktrttrpkrcttsamprtp	134		
QY 1025	SRIEAMQTTTRPQOTNSKLVENPSPSEDGAEGEPTNHLRLRHVMEPTSPMDLT	1084		
Db				
Db	s-lpeamqttrtrprrpsneldvnsenedgaagekprmltfrrpvtlrlvlgteitl	193		
QY 1085	PRVPMOGIILNPMSEDENICRGKRVYDGLTTRNCTIVAFRGNHYWMLSPSPSPARK	1144		
Db				
Db	vrgpsqgslrmpmfedcnlcngrvvdlttlngllvalfrhywmltrfrtpprrpt	253		
QY 1145	TEWNGISPRITDTVTRFCNCEGKTFPFKDSQVWFTNDIKDAGYPRKTEFGGLTGQIVA	1204		

Db 254 tewwpsipdvtvfrncceqgtffksgsgywrftndlkdaegypklsgfgjngkiva 313  
 Oy 1205 ALSTAKYKNMPESSVYFFRRGSGIDQYIKQEPVQKCPGRPALNPPYGEWTVRRRFE 1264  
 Db 314 alslagysrpsesvyffkrrgsvqgytykgeplckctgrpalnysygetaqrtrrfie 373  
 Oy 1265 RAIGSSQHTHTTRIOXSPARLAKODKGVLANEKKVSIIMRGLPNVNTSAISLPNIRKPDGY 1324  
 Db 374 raigsgvhttrihytpvrvpydgkflhnekvksltwrgjpnvntsaisslpnirkpdy 433  
 Oy 1325 DYAFSKDQYNIIDVPSRTARITTRSGOTLSKMWYNCP 1363  
 Db 434 dyalskdgynidvpsrtaratrittrsgqlsnclwncp 472  
 RESULT 8  
 AAM24516  
 ID AAM24516 standard; Protein; 5179 AA.  
 XX  
 AC AAM24516;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE C899P predicted amino acid sequence.  
 XX  
 KW Human: immunotherapy; diagnosis; colon cancer; colon tumour;  
 KW immunogenic; gene therapy; vaccine; colonic cancer.  
 OS Homo sapiens.  
 XX  
 PN WC200149716-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 29-DEC-2000; 2000WO-US35596.  
 XX  
 PR 30-DEC-1999; 99US-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jiang Y;  
 DR WPI; 2001-441847/47.  
 XX  
 PT  
 PT  
 PT  
 XX  
 XX  
 PS Claim 2; Page 446-462; 472pp; English.  
 CC The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytosratic activity.  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of restorative vaccine. (I) may

[illegible]

Db	601	apcapkeapcttprkeapcttckl1crctpeklapctprapctprcealapctpeapctt	660
Qy	620	peebpattprkaapantprkebabttprkebaattprkebaattprkebaattprkstaatttke	679
Db	661	peebapctprkaapantprkeapcttprkeapcttprkeapcttprkeapcttprkeapcttprke	720
Qy	680	aprtpprkprakelaprttprkeprsttsdkaaprttprkg7aprtprkebabttprkeaprtprk	739
Db	721	aprtpprkprakelaprttprkeprsttsdkaaprttprkg7aprtprkebabttprkeaprtprk	780
Qy	740	tabtttkebabttprkprakelaprttprkstaattprkebaattprkebaattprkebaattprk	799
Db	781	tabtttkebabttprkprakelaprttprkstaattprkebaattprkebaattprkebaattprk	840
Qy	800	kaprtprerpprtsevtprttprkeprttihksprdestprelsaeptrpalesprkeprcvpr	859
Db	841	kaprtprerpprtsevtprttprkeprttihksprdestprelsaeptrpalesprkeprcvpr	900
Qy	860	tktpatprkemttprkdkktprbedlaptprbetttaaprkmtkeatatttkttskittatttqv	919
Db	901	tktpatprkemttprkdkktprbedlaptprbetttaaprkmtkeatatttkttskittatttqv	960
Qy	920	tsttttdtprfkittllkttttllaprkvtttkttitttelnkprteaprkdratnskaattpr	979
Db	961	tsttttdtprfkittllkttttllaprkvtttkttitttelnkprteaprkdratnskaattpr	1020
Qy	980	pokprtkaprkkprstkkprtkmrvrkkrktpprprkmtstmpelnftrsiabamlotttrrn	1039
Db	1021	pokprtkaprkkprstkkprtkmrvrkkrktpprprkmtstmpelnftrsiabamlotttrrn	1080
Qy	1040	qprnsklvynvknrsedagagagertpnhmlrphvmpavttdmqlprvprnqgittipmls	1099
Db	1081	qprnsklvynvknrsedagagagertpnhmlrphvmpavttdmqlprvprnqgittipmls	1140
Qy	1100	detnrcngkprvdglttlrngtllvafrcghyfwmlspsfppsparrtiewgipspidvft	1159
Db	1141	detnrcngkprvdglttlrngtllvafrcghyfwmlspsfppsparrtiewgipspidvft	1200
Qy	1160	rcncgsktfffkdsqymrfetndidacvprkpfmgfegglgqlvaalstakrkmpesvy	1219
Db	1201	rcncgsktfffkdsqymrfetndidacvprkpfmgfegglgqlvaalstakrkmpesvy	1260
Qy	1220	ffkrngstlqqyiykoepvokcprgrrpalnpyvge	1254
Db	1261	ffkrngstlqqyiykoepvokcprgrrpalnpyvge	1295
RESULT 5			
AAB29778			
ID	AAB29778	standard; Protein: 902 AA.	
XX	AC	AAB29778;	
XX	DDT	28-FEB-2001 (first entry)	
DE	XX	Human MSF-derived tribonectin.	
XX	XX	Human tribonectin; MSF: megakaryocyte stimulating factor.	
KW	XX	alternative splicing; joint boundary lubricant; O-linked oligosaccharide;	
KW	XX	osteoarthritis; tribosupplementation; tissue adhesion inhibition;	
KW	XX	fraction coefficient reduction; gene therapy; antiarthritic;	
OS	XX	osteopontin.	
OS	XX	Homo sapiens.	
XX	PN	MO200064930-A2.	
XX	PD	02-NOV-2000.	
XX	PR	24-APR-2000; 2000MO-US10953.	
XX	PR	23-APR-1999; 99US-0298970.	

Query Match	Best Local Similarity	Score	DB	Length
Matches 794; Conservative	28; Mismatches 80; Indels 162; Gaps 66	47.9%; 74.6%;	3484; 22;	902;
159	VADNKKNRKPKPRPVNDVAGSGLDNGDKRVTPPTDSTQHNKVSVPKITTAKPIN	1	VADNKKNRKPKPRPVNDVAGSGLDNGDKRVTPPTDSTQHNKVSVPKITTAKPIN	218
219	PRPSLPSPSDSKESILVNVNKETVETKETTNNQTSIDGKEKTSKAKETOSIEKTSKAK	61	PRPSLPSPSDSKESILVNVNKETVETKETTNNQTSIDGKEKTSKAKETOSIEKTSKAK	278
279	DLAPTSKVLAKPTPKAETTTKGPAITTTKEEPTTPPKKPASTTPKEEPTTTKSAPTTPK	121	DLAPTSKVLAKPTPKAETTTKGPAITTTKEEPTTPPKKPASTTPKEEPTTTKSAPTTPK	338
339	EPAPITTSAPPTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	173	EPAPITTSAPPTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	398
399	PKKAPATTPKBPAPTTKEEPTTPPTPKKEAPTTKEEPAAPTTKEEPAAPTTAKKAPATTPK	229	PKKAPATTPKBPAPTTKEEPTTPPTPKKEAPTTKEEPAAPTTKEEPAAPTTAKKAPATTPK	458
459	PTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	282	PTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	518
519	PAPPTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	336	PAPPTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTTTKEEPAAPTT	578
579	TPKKLITPTTPEKLAAPTTPEKAPATTPPEBLAATTPPEEPTPTTPEEPAATTPKAAAPNTPK	389	TPKKLITPTTPEKLAAPTTPEKAPATTPPEBLAATTPPEEPTPTTPEEPAATTPKAAAPNTPK	638

|||||  
Db 721 aptlpkppakelaplttkepstsdskpaptlpkgtaptlpkeaptpkcapltppk 780  
Qy 740 TAFTTLKEPAPTTPKKPAKRELAFTTGTGPTSTSDKAPATTPPKKELAPTTKEPAPTTPK 799  
Db 781 tapltlkepaptkppakelaplttkgptstsdskpaptlpkcapltppkeaptpk 840  
Qy 800 KPAPEPTPEPTSEVSTPTTKEPTTIHKSPESTPELSAEPPTKALENSPKEGVP 859  
Db 841 kpaepcppepplteisevpttkepttlhkspesteplsaeptpkalenspkpvp 900  
Qy 860 TKTPAATKPEMTTAAKDTTERDLRTPEPTTAAPKMTKETAATTEKTESKTTATTQY 919  
Db 901 tktpaatkpemttaakdtkterdlrtpepttaapkmktetatekteskttattqy 960  
Qy 920 TSTTTQDTTFFKTTTLTKTTLAKVYTKTKTTTTEIMKPEETAKPKDRTSKATTPK 979  
Db 961 tstitqdtltpfkltlktllapkvtktkctltteimkpeetapkdatatskattpk 1020  
Qy 980 POKPTAKPKPTSTKKPKTMPRVKRPKPTTPPKMTSTMBELNPTSRIAPAMLTQTTTPN 1039  
Db 1021 poptkapkptsttkkpktmprvrkpttpkmtstmbelnptsrilaamltqtltpn 1080  
Qy 1040 QTENSKLVEVNPKESEDAGAEGETPHMLLRPHVEMPEVTPDMQYLPRVPOGIIINPMLS 1099  
Db 1081 qtpnslkvevnpksedagaegelpmllrphvfmpevtpdmqylprvpogliinpmls 1140  
Qy 1100 DEFNICKGRPVVDGLTTLRNGLTVAFRGHVFWMLSPSPSPAPARTIEWGICPSPIDTFT 1159  
Db 1141 dehnickgrkvvdglttlrngltvafrghyfwmlspfspapartiltevwgicpspidft 1200  
Qy 1160 RCNCEGTEFFPKDSQYWRFTNDIKDAGYPRKRGFGGLGQYVAAISTAKYKNWPSVY 1219  
Db 1201 rcncegtrffkdsqywrftndikdagypkrlkgilgqyvaalstakynwpsvy 1260  
Qy 1220 FFKRGSGIOQYIYKOEVOVOCOPRRPALNYPVGEVTOVRRRPERAIGSQHTTIRIQY 1279  
Db 1261 ffrkgsgioqylykqevpqqcprrpalnypvgemqvvrrrrferaisgsqhtlriqy 1320  
Qy 1280 SPARLAVODKGVLEHNEKVSILIMRGLPNVYTSALISLPIKRPDGYDYAASKOQYINIDV 1339  
Db 1321 sparlaydkgvlnhevksllwrglpnvysaalslpiirkpdydyatskdyynldv 1380  
Qy 1340 PSRTARAITTSSGOTLSKWNCP 1363  
Db 1381 psrtaraaitrsgqlskwnp 1404  
RESULT 4  
AAM24322  
ID AAM24322 standard; Protein: 1299 AA.  
AC AAM24322:  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST encoded protein SEQ ID NO: 1847.  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HXSEQ INC.  
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Dermanac RA, Zhang J, Werhman T;  
XX WPI: 2001-476164/51.  
XX N-PSDB: AAH98981.  
DR  
XX  
PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
PS Claim 20; Page 1198-1201; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
SQ Sequence 1299 AA;

Query Match 91.7%; Score 6669.5; DB 22; Length 1299;  
Best Local Similarity 96.8%; Pred. No. 0;  
Matches 1254; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

1 MAMKTLPIYLLLSVFIQOVSSQ-----  
Db 1 mawktlpiylylllsfvqivqvsqqlsacagrcgysrdatcncdqhyameccpdf 60  
Qy 26 -----ELSCGKGFESFERGRCDDAQCKYDKCCPDYESCAVHNPTSPSSKKAP 79  
Db 61 kvctaeliscgkgrcfesfergrecddaqckkydkccpdyesfcaevhnptspsskkap 120  
Qy 80 PRGASQSTITSSTKRSPKPNKKTKKVTSEETHEHSVENQESSSSSSSSSSSTIWM 139  
Db 121 prgsaqstltsktrspkpnkktkkvtseethehsvenqessssssssstllw 180  
Qy 140 KIKSSKNANRELOKKLKVKNKNNKTKKTPKPPVYDEAGSGLDNGDFKVTTPDTST 199  
Db 181 kiksksnaanrelgkklkvknnkntkktpkppvydeagsgldngdfkvttpdst 240  
Qy 200 TQHNKVTSTKTIITAKLINRPSLPPNSDTSKETSILVNETVETKETTNTKOTSDQ 259  
Db 241 tqhnkvstspkltitaklpinrpslppnsdtsketsilvnetvetkettntkqstldq 300  
Qy 260 KEKTSKAKETQSTIEKTSAKDLAPTSKVLAFTPKAETTTGPAULTPKKEPTTPKREPAS 319  
Db 301 kektsaketsqstiektakdlaptskvlakpkaetttgpaaltpkkepttpkrepas 360  
Qy 320 TTPEKEPTTIKSAPTTPKKEPAPTTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379  
Db 361 ttpekepttiksaptpkkepaptttsapttkepaptttkepaptttkepaptttkepap 420  
Qy 380 APTTTSAPTTTPKREAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPK 439  
Db 421 aptttsaptttpkreapattppkkapattppkkapattppkkapattppkkapattppk 480  
Qy 440 EPAPTAPEKPPAPTPPKPEAPATTPPKPEAPATTPPKPEAPATTPPKPEAPATTPPKPEAP 499  
Db 481 epaptapekppaptppkpeapattppkpeapattppkpeapattppkpeapattppkpeap 540  
Qy 500 TTKSAPTTPKESPPTTKEPAPTTPKPEAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPK 559  
Db 541 ttksaptpkpspttkepapttpkpeapattppkkapattppkkapattppkkapattppk 600  
Qy 560 APTAPKEPAPTTPKETATTPTKKLTPPTPEKLAFTTPEKAPATTPPEELAPTPPEEPTPTT 619



QY 1220 FEKRGSIQOYITKQEPVQKCGRRPALNVPYGEVTRRRRRERRAIGPQSTHTIRIQY 1279  
 |||||  
 DB 1261 ftkrgsldqylykqepqkcpgrrrpalnpyvgemtqyrrrrrrferraiqpsqthliriy 1320  
 QY 1280 SPARLAQDQKGLHNEVYSILMRGLPNVNTSAISLPNRRKPDGDYAFSDQYVNIIDV 1339  
 |||||  
 DB 1321 sparlqyqdkglhnevksllwrglpmvtsalslpmllrpdgydyafskdyynldv 1380  
 QY 1340 PSRTARAITRRSGQTLKRVWNCIP 1363  
 |||||  
 DB 1381 psrtaraitrrsgqlskrvwncip 1404  
 RESULT 2  
 AAB60568  
 ID AAB60568 standard; Protein; 1404 AA.  
 XX AAB60568;  
 AC AAB60568;  
 XX AAB60568;  
 XX AAB60568;  
 DE 27-APR-2001 (first entry)  
 XX Human megakaryocyte stimulating factor (MSF, CACP).  
 XX Human; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;  
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
 KM chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KW antiahrthritic.  
 XX  
 OS Homo sapiens.  
 PN WO200107066-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 21-JUL-2000; 2000WO-US20002.  
 XX  
 PR 23-JUL-1999; 99US-0145328.  
 PR 19-JUL-2000; 2000US-0145328.  
 XX  
 PA (UYCA-) UNIT CASE WESTERN RESERVE.  
 XX  
 PI Weiman ML;  
 XX  
 DR WPI; 2001-182721/18.  
 XX  
 PT New composition comprising the campodactylly-arthropathy-coxa  
 PT vara-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints  
 XX  
 PS Example 1; Page -; 34pp; English.  
 XX  
 XX The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactylly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactylly-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U70316).  
 XX  
 XX Sequence 1404 AA;  
 SQ

Query Match 99.6%; Score 7245.5; DB 22; Length 1404;  
 Best Local Similarity 97.1%; Pred. No. 0;  
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;  
 1 MAMKTLPIYLLLLSVFIQGVSSQ----- 25  
 |||||  
 DB 1 mawktlpiy111111svf1qgvssqdlsscagrcgysrdatcncdqhyameccpdf 60  
 QY 26 -----ELSCGRCPESERERECDDCAQCKKVKCCPDYVSCFAEYHNPSPPSSKKAP 79  
 |||||  
 DB 61 kvctaelckgrcfesfergrecddaqckkkykccpdyescfaeympspsskkap 120  
 QY 80 PPSGASQITKSTTRKSPKPKKKTKKVVIESEETEEHSHVSENOESSSSSSSSSSSTIW 139  
 |||||  
 DB 121 ppsgasqlstkrspkpkkktkkviesselteesvseugessssssssstliw 180  
 QY 140 KIKSSKNSAANRELOKLLKVDNKNKRTKKRPKPPVVDAGSGLDNGDKVTTPTST 199  
 |||||  
 DB 181 kiksksaanelelqkllkvdknkrkkpkrpvyvdeagsgldngdkvttptst 240  
 QY 200 TOHNVSTSPKTTAKPINPNSIPNSDTSKETSIVNKEVETKETTNNKQTSNDG 259  
 |||||  
 DB 241 tqhkvstspkltakpinpipsipnsdtsketslvnketvettltnqstldg 300  
 QY 260 KEKTSAKETOSIEKTSAKDLAPTSKVLAKPDKAETTTKGPALTTPKEPTTPKPPAS 319  
 |||||  
 DB 301 kektsaketgtsaktsakdlaptskvlakpdkpaetltkgpalttpkepttpkppas 360  
 QY 320 TTPKEPTPTTKSAPTTKEPAPTTTPKSAPTTTPKKEPAPTTTPKKEPAPTTTPK 379  
 |||||  
 DB 361 ttpkeptpttksapttkepaaptttkesaptttkesaptttkesaptttkesaptttk 420  
 QY 380 APPTTKSAPTTTPKKEPAPTTTPKKAAPTTPKKEPAPTTTPKKEPAPTTTPK 439  
 |||||  
 DB 421 apttkesaptttkesaptttkesaptttkesaptttkesaptttkesaptttk 480  
 QY 440 EPAPTAAPKRPADTPPKKEPAPTTTPKKEPSTTPKKEPAPTTTPKSAPTTTPK 499  
 |||||  
 DB 481 epaptapkkpapttkesaptttkesaptttkesaptttkesaptttkesaptttk 540  
 QY 500 TTKSAPTTTPKKEPSTTPKKEPAPTTTPKKAAPTTPKKEPAPTTTPKKEPAPTTTPK 559  
 |||||  
 DB 541 ttksaptttkesaptttkesaptttkesaptttkesaptttkesaptttkesaptttk 600  
 QY 560 APAPKEPAPPTTPKKEPAPTTTPKKAAPTTPKKEPAPTTTPKKEPAPTTTPK 619  
 |||||  
 DB 601 apapkepappttpkkesaptttkesaptttkesaptttkesaptttkesaptttk 660  
 QY 620 PEEPAPPTPKKAAPNTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPK 679  
 |||||  
 DB 661 peepapptpkkaapnptkkesaptttkesaptttkesaptttkesaptttkesaptttk 720  
 QY 680 APPTPKKPAPELAPPTTTPKKEPSTTSDKAPPTTPKSTAPPTTPKKEPAPTTTPK 739  
 |||||  
 DB 721 aptpkpkpapelappttpkkesaptttkesaptttkesaptttkesaptttkesaptttk 780  
 QY 740 TADPTLKEPAPTPPKKPAPELAPPTTTPKKEPSTTSDKAPPTTPKSTAPPTTPK 799  
 |||||  
 DB 781 tadptlkepaptppkkpaapelappttpkkesaptttkesaptttkesaptttkesaptttk 840  
 QY 800 KKPAPTPEPPTTPKKEPSTTTPKKEPSTTTPKKEPSTTTPKKEPSTTTPKKEPSTTTPK 859  
 |||||  
 DB 841 kkpaptpeppttpkkesaptttkesaptttkesaptttkesaptttkesaptttkesaptttk 900  
 QY 860 TKTPPAAPKPEMTTAKDKTERDLRTTPTTAAKPKETATTTKSTTATTTT 919  
 |||||  
 DB 901 tktpaapkpemttakdkterdlrttptttaaakpketattttksttatttt 960  
 QY 920 TSTTQDTTPPKKTTTLKTTTLAKKVTTKTTTTEIMNPEETAPKDKATSKATTPK 979  
 |||||  
 DB 961 tsttqdttpkktttllktttlakkvtttkttttemnpeetapkdakatstkattpk 1020



```

FT      Region      1265..1331
FT      Region      /label- Exon_X
FT      Region      1331..1373
FT      Region      /label- Exon_XI
FT      Region      1373..1404
FT      Region      /label- Exon_XII
XX      MO9213075-A.
XX      06-AUG-1992.
XX      17-JAN-1992: 92MO-USO0433.
XX      18-JAN-1991: 91US-0643502.
XX      10-SEP-1991: 91US-0757022.
XX      (GEMV ) GENETICS INST INC.
XX      Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX      WPI: 1992-284660/34.
XX      N-PSDB: AAQ27223.
XX      New human mega-karyocyte stimulating factors - for treating
XX      immune deficiencies, cancer, exposure to radiation or drugs,
XX      bacterial and viral infections, etc.
XX      Claim 1, 2 and 3; Fig 1; 87pp; English.
XX      The sequence given is a full length translation from the megakaryocyte
XX      stimulating factor (MSF) precursor. The sequence covered by exons II,
XX      III and IV encodes megakaryocyte stimulating factor (MSF). This
XX      sequence is modified by the addition of an N-terminal sequence encoding
XX      a secretory leader, an initiating methionine preceding exon II and a
XX      terminating codon following exon IV. The CDNA sequence given contains
XX      (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX      classical mammalian protein secretion signal sequence. The sequence
XX      encoding the original meg-CSF includes exons II-IV and is thought to
XX      terminate in the region between amino acid residues 134 - 147. The
XX      yield a family of mRNAs each encoding a different MSF protein. Exons
XX      V and VI are thought to be related to the activity of the factor and
XX      are also implicated in the stability, folding and processing of the
XX      molecule. These exons are also thought to play a role in the observed
XX      synergy of MSF with other cytokines. Exons V - XII are believed to be
XX      implicated in the processing or folding of the appropriate structure of
XX      the resulting factor, ie. one or more of these exons may contain
XX      sequences which direct proteolytic cleavage, adhesion, organisation of
XX      the cellular matrix or extracellular matrix processing. Both naturally
XX      occurring and non-naturally occurring MSF's may be characterised by
XX      various combinations of alternatively spliced exons from this sequence,
XX      with the exons spliced together in differing orders to form different
XX      members of the MSF family.
XX      Sequence 1404 AA:
XX      Query Match 99.68; Score 7245.5; DB 13; Length 1404;
XX      Best Local Similarity 97.18; Pred. No. 0;
XX      Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY      1 MAWKTLPIYLLLSVFIQVSSQ-----
DB      1 MAWKTLPIYLLLSVFIQVSSQ-----
QY      26 -----ELSGKRCFESEFERGECDCDAQCKRYDKCCPDYSEFCAEVNHTSPSSKKAP 79
DB      61 KYRCELESCGKRCFESEFERGECDCDAQCKRYDKCCPDYSEFCAEVNHTSPSSKKAP 120
QY      80 PPSGASQITKSTTKRSPKPPNKKTKRKVIESBEITEEHSYSVSNQSSSSSSSSSSSTIIV 139
DB      121 PPSGASQITKSTTKRSPKPPNKKTKRKVIESBEITEEHSYSVSNQSSSSSSSSSSSTIIV 180

```

```

QY      140 KIKSSKNSAANRELQKTKLVKDNKKNRKTKKTPKPPVVDAGSGLDNGDFKVTTPPTST 199
DB      181 KIKSSKNSAANRELQKTKLVKDNKKNRKTKKTPKPPVVDAGSGLDNGDFKVTTPPTST 240
QY      200 TQHNKVSSTSPKLTTPAKPIINPRESLPPNSDTSKENSILVANKETTYETKETTNNKQSTSDG 239
DB      241 TQHNKVSSTSPKLTTPAKPIINPRESLPPNSDTSKENSILVANKETTYETKETTNNKQSTSDG 300
QY      260 KEKTSARETOSIEKTSKDLAPTSSKVLAKPPKAEETTTKGPALVTTPEPTTPPKPPAS 319
DB      301 KEKTSARETOSIEKTSKDLAPTSSKVLAKPPKAEETTTKGPALVTTPEPTTPPKPPAS 360
QY      320 TTPKEPTPTTIKSAPTTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBP 379
DB      361 TTPKEPTPTTIKSAPTTPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBP 420
QY      380 APPTTKSAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPK 439
DB      421 APPTTKSAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPK 480
QY      440 EPAPTKAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAP 499
DB      481 EPAPTKAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAP 540
QY      500 TTKSAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBP 559
DB      541 TTKSAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBP 600
QY      560 APAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPK 619
DB      601 APAPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPK 660
QY      620 PEPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAP 679
DB      661 PEPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAP 720
QY      680 APPTPKKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 739
DB      721 APPTPKKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 780
QY      740 TAPPTKEAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 799
DB      781 TAPPTKEAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTPPKBPAPTP 840
QY      800 KPAPTPKEPTPTSEVSTPTTKKEPTTIHKSPOBSTPELSAEPTRPALENSPKEPCVPT 859
DB      841 KPAPTPKEPTPTSEVSTPTTKKEPTTIHKSPOBSTPELSAEPTRPALENSPKEPCVPT 900
QY      860 TTPPATPEMTTAKDRTTERDRLRTTPETTTAARPKMTKETATTEKTSKITTATTTTOY 919
DB      901 TTPPATPEMTTAKDRTTERDRLRTTPETTTAARPKMTKETATTEKTSKITTATTTTOY 960
QY      920 TSTTODTTPKTIITLTKTTTIAKPVTTTAKTITTTETIMNRPBSTAKPKDRTNKAATTPK 979
DB      961 TSTTODTTPKTIITLTKTTTIAKPVTTTAKTITTTETIMNRPBSTAKPKDRTNKAATTPK 1020
QY      980 POKPTPAKPKKSTTKKPKTPRVKPKPTTPPKMTSTIMELMPTSRIAEAMQTTTRPN 1039
DB      1021 POKPTPAKPKKSTTKKPKTPRVKPKPTTPPKMTSTIMELMPTSRIAEAMQTTTRPN 1080
QY      1040 QTPNSKLVEVNPKSSEDGAGGEGSTPHMLLRPHVFMKEVTPDMDYLRVFNOCGIINPMLS 1099
DB      1081 QTPNSKLVEVNPKSSEDGAGGEGSTPHMLLRPHVFMKEVTPDMDYLRVFNOCGIINPMLS 1140
QY      1100 DETNINCKRPVDLITLLRNQTLVAFRGHYFWMISPPSPSPARRTTEWGISPIDVFT 1159
DB      1141 DETNINCKRPVDLITLLRNQTLVAFRGHYFWMISPPSPSPARRTTEWGISPIDVFT 1200
QY      1160 RCMCEKCTFFFKNSQWRNTNDIKDAGYKPIFKGFGGLTGOIVALSTAKYKKNPEVY 1219
DB      1201 RCMCEKCTFFFKNSQWRNTNDIKDAGYKPIFKGFGGLTGOIVALSTAKYKKNPEVY 1260

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:24:09 ; Search time 138.84 seconds

(without alignments)  
727.182 Million cell updates/sec

Title: AA2  
Perfect score: 7276  
Sequence: 1 MAMKTLPIYLLLSFVIO.....ARATTSQGLSKVWVNCNP 1363

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_1101.\*  
2: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7245.5	99.6	1404	13	MSF precursor. Sy
2	7245.5	99.6	1404	22	Human megakaryocyt
3	7245.5	99.6	1404	22	Human megakaryocyt
4	6669.5	91.7	1299	22	Human EST encoded
5	3484	47.9	902	22	Human MSF-derived
6	1916.5	26.3	452	16	Human megakaryocyt
7	1707.5	23.5	472	22	Bovine MSF ortholo
8	1175	16.1	5179	22	C899P predicted am
9	981	13.5	763	21	Arabidopsis thalia
10	950	13.1	1664	19	C. thermocellum O1
11	805	11.1	1325	22	Peptide #2327 enco

12	715	9.8	763	18	AAW31852
13	645.5	8.9	4412	21	AAV53666
14	612	8.4	572	18	AAW31855
15	555.5	7.6	844	7	AAW60570
16	542	7.4	807	21	AAV54467
17	520.5	7.2	788	21	AAV54466
18	515	7.1	1837	21	AAH11726
19	507	7.0	182	12	AAH10872
20	506.5	7.0	744	9	AAH82975
21	498.5	6.9	1721	21	AAH11727
22	498.5	6.8	1721	19	AAH48299
23	496.5	6.8	2971	21	AAH41231
24	489.5	6.7	2972	22	AAH50363
25	489.5	6.7	3118	22	AAH50362
26	488.5	6.7	826	13	AAH26042
27	488	6.7	617	22	AAH16458
28	488	6.7	617	22	AAH04187
29	488	6.7	957	21	AAV59288
30	488	6.7	957	22	AAH24513
31	485	6.7	1127	22	AAH95541
32	476.5	6.5	652	9	AAH82974
33	467.5	6.4	511	22	AAH14883
34	467.5	6.4	511	22	AAH07312
35	467.5	6.4	511	22	AAH02607
36	454.5	6.2	1012	20	AAV17406
37	454	6.2	2819	22	AAH35408
38	450.5	6.2	378	12	AAH14160
39	448	6.2	1237	21	AAH81609
40	447.5	6.2	3266	21	AAH42491
41	446.5	6.1	378	12	AAH14162
42	446.5	6.1	750	20	AAV05477
43	446.5	6.1	2665	22	AAH14533
44	446.5	6.1	2665	22	AAH26950
45	446.5	6.1	2665	22	AAH02259

## ALIGNMENTS

RESULT 1	AAH26049 standard; Protein; 1404 AA.
XX AC	AAH26049;
XX AC	02-FEB-1993 (first entry)
XX AC	MSF precursor.
XX AC	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
XX AC	stability; proteolytic cleavage; adhesion; alternative splicing.
XX AC	Synthetic.
XX AC	Location/Qualifiers
XX AC	1..26
XX AC	/label= Exon_I
XX AC	26..67
XX AC	/label= Exon_II
XX AC	67..107
XX AC	/label= Exon_III
XX AC	107..157
XX AC	/label= Exon_IV
XX AC	157..200
XX AC	/label= Exon_V
XX AC	200..1141
XX AC	/label= Exon_VI
XX AC	1411..1166
XX AC	/label= Exon_VII
XX AC	1166..1212
XX AC	/label= Exon_VIII
XX AC	1213..1266
XX AC	/label= Exon_IX

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the Fa  
Amino acid sequenc  
Cryptosporidium pa  
Protein encoded by  
Bioadhesive precur  
Portion of Cryptos  
Cryptosporidium pa  
Human ORFX ORF995  
Human SRCAP. Homo  
P. yoe111 SSP2 ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 polype  
C900P predicted am  
Human protein sequ  
Bioadhesive precur  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
Human atrophin-1 r  
Human 07CG27 gene  
PRP 378. Trilicium  
Streptococcus pneu  
Human ORFX ORF2255  
PRP encoded by c10  
C. albicans Rb1 p  
Peptide #967 enco  
Peptide #987 enco  
Peptide #941 enco

1-3



Db 811 epsttqe--hadekkasegdnislsrleeteek 842

Search completed: April 26, 2002, 16:24:52  
Job time: 482 sec

```

OY 919 -RKMTSMPELNPTSRIAEAMLOTIT 943
      : : | : | : | : | : | : | : |
DB 4258 IKDIIltesefvgsaaifecvlspsrt 4283

RESULT 14
AAW31855
ID AAW31855 standard; Protein; 572 AA.
XX
AC AAW31855;
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 55 kDa protein.
XX
KM Tuberculosis; mycobacteria; infection; diagnosis;
XX antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN MO9741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
PI Bspitica C, Honisch C, Moreno C, Singh M;
XX
DR WPI; 1997-549750/50.
XX
DR N-PSDB; AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
PT used for diagnosis of mycobacterial infections, monitoring
PT vaccination and development of anti-mycobacterial agents
XX
PS Claim 11; Fig 16; 55pp; English.
XX
CC This novel 55 kDa protein is encoded by an open reading frame of
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
SQ Sequence 572 AA;

Query Match 9.0%; Score 612; DB 18; Length 572;
Best Local Similarity 30.6%; Pred. No. 5.5e-29;
Matches 193; Conservative 29; Mismatches 253; Indels 156; Gaps 33;

```

```

OY 361 -----PKEP-----PTTPKEPATTKEPSPTKEPAPTTTKSAPTTTKEPAPTT 407
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 166 evlaaialpsrpsrplltttipalpapl-----pplpplpplpntavppl-----pplpvp 217

OY 408 TKSAPTTPK-EPSPTTTKEP--APTTPKEPATTTPKKAPTTTPKEPAPTTTKEPAPTTTK 464
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 218 talappilpplaplpispvypapaplpipgkpwltlpplapapapkp-tvpvlpvpscpse 276

OY 465 KPAPTPAPKEPAPTTTKEPAPTTTPKKLTPPTPE--KlapTPKEPAPTTPEELAPTTPEEP 522
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 277 knpppap--peppeksspalppapapapsavvvpssplppap-----paapras 327

OY 523 TPTEEPAPTTTPKA---AAPTPKEPAPTTTPKEPAPTTTPKEPAPTTPK---ETAPTTPK 576
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 328 mpalp--papppapalrllpplpssqapnsp--papap-----plpckllsaapcpvp 379

OY 577 GTAPTTLKEPAPTTTPKKAPKELAPTTTEPSTSDKAPTTTPKGTAPTTTPKEPAPTTTP 636
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 380 ppapn--rppapap--pappel-----papdp-----plpvvauspap 416

OY 637 KEPAPTTTPGTAPTTLKEPAPTTTPKKAPKELAPTTTKGP-----TSTSDKAPATTTP 689
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 417 --papappsalp-tvpappptpaapksrpalpaapppapppvratpppapppap 473

OY 690 KETA--PTTPKEPAP--TTPKKPAPTTPEPTTPPTSEVSTPTTKEPTTIKSPDESTP 744
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 474 nsmaelpapppdppllpalppapppaplpmsppaplpipaapdpapapltltnqp-----sp 530

OY 745 ELSAEPTPKALENSPKKEGVTTKTPATKTP 775
      | : | | | | | | | | | | | | | | | | | | | | | |
DB 531 pla-----pvpapilaplpingrp 549

RESULT 15
AAP60570
ID AAP60570 standard; Protein; 844 AA.
XX
AC AAP60570;
XX
DT 24-AUG-1991 (first entry)
XX
DE Sequence of the Falci-parum Interspersed Repeat Antigen
XX (FIRA).
XX
KM Malaria vaccine; antigen; epitope.
XX
OS Plasmodium falciparum.
XX
PN WC8601802-A.
XX
PD 27-MAR-1986.
XX
PF 11-SEP-1985; 85WO-0006960.
XX
PR 11-SEP-1984; 84AU-0007067.
PR 11-SEP-1984; 84AU-0007066.
PR 10-SEP-1985; 85AU-0047326.
XX
PA (HALL-) HALL INST MED RES.
XX
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
XX WPI; 1986-094065/14.
XX N-PSDB; AAN60473.
XX
PT DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum
XX
PS Disclosure; Fig 7; 55pp; English.
XX
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base

```

```

Db      638 paapapapvratcp-----ppapapppn 665
QY      880 SKATPRPKP-----TKAKKPTSTIKKPKTMVRKPKTTPPRKMTSTM-----PELNP 930
Db      666 smalppppdpplplacppappplpmppapplppappppppllnppppplap 725
QY      931 TSRIAEAMQTTTRPNQTPNSKIVEVNPKESEDAGAGEET 970
Db      726 vpgaplaplplngprvafarknsl-----gassgat 756

RESULT 13
AAV53666
ID      AAV53666 standard; Protein: 4412 AA.
XX
AC      AAV53666;
XX
DT      22-FEB-2000 (first entry)
XX
DE      Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.
XX
KW      Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KW      bone development; g1/1017427/emb/CAA62189.
XX
OS      Unidentified.
XX
PN      WO9960164-A1.
XX
PD      25-NOV-1999.
XX
PE      14-MAY-1999; 99WO-US11066.
XX
PR      15-MAY-1998; 98US-0085673.
XX
PA      (OUAR-) OUARK BIOTECH INC.
XX
PI      Elhat P, Mor O, Skalter R, Feinstein E, Faerman A;
XX
DR      WPI: 2000-053304/04.
XX
PT      Identification of stress induced genes for determining risk and
PT      preventing, treating or controlling osteoporosis
XX
PS      Claim 32; Fig 6A-R; 308bp; English.
XX
CC      The present sequence is obtained from a clustral X alignment with
CC      protein 608. Protein 608 was identified using the method of the invention
CC      after subjecting rat osteoblasts to mechanical stress. Expression of the
CC      608 gene was found to be upregulated by about 3-fold in cells subjected
CC      to mechanical strain. The specification describes a method for the
CC      identification of genes responsive to a specific mechanical stress. The
CC      method comprises applying the mechanical stress to an organism (tissue
CC      or cells comprising bone cells), isolating the specific cellular
CC      fractions and extracting mRNA from them, and differentially analysing the
CC      mRNA in comparison with control samples. The method is used to identify
CC      genes whose expression is responsive to a specific stress. The identified
CC      genes are employed in determining risk associated with a physiological or
CC      disease state. The risk determination methods are used for testing a
CC      medicament for gene therapy. These medicaments, or genes identified by
CC      the method of the invention, are used for treating, preventing or
CC      controlling a physiological or disease state (especially osteoporosis or
CC      bone density or other factors causing or contributing to osteoporosis or
CC      its symptoms or other conditions involved in mechanical stress or its
CC      lack. The methods can also be used for advancing research or studies in
CC      bone development.
XX
SQ      Sequence 4412 AA:

```

```

Query Match      9.28; Score 627; DB 21; Length 4412;
Best Local Similarity 24.18; Pred. No. 6.9e-29;
Matches 266; Conservative 110; Mismatches 442; Indels 288; Gaps 51;

```

```

QY      26 EISCKGRCFES-----FERGRCDCDAQCKKYKCCPDYESFCA--EVKDNKK----- 71
Db      3278 qlschvgseprlqlwklage-----lkpdcresfsfagtlavlelrdvakadsdyv 3331
QY      72 -----NRTKKKPTPK-PPVUDEA--GSGLDNGDKVYTPDTSTTHQHNKSTSPKIT 119
Db      3332 cksanvgsdtkskvltkdkpavapakckaavdgrlffvsepsqlrvvektatflakv 3391
QY      120 TAKPINRPSLPPNSDTSKETSLSLVNK-----ETTVETKETTTN----- 159
Db      3392 ggdprl-----pnvktvktgvrqlngggvrvflngqdeaklelrdtclsdgjlrcvaf 3444
QY      160 -----KQSTDGKEKETSAAKETOSIEKTSAA-----KDLAPT 190
Db      3445 nehgeiesnvlqvdarkkgekieg-dlramlkktrpllkkgageeeldmellknvdpk 3503
QY      191 S-----KVLAPTPKAEETTTGPA 209
Db      3504 eyekyarmygltdfrglqrvveehrvkvhvlefeaevefekp-----kappkgye 3559
QY      210 LT-----TPKEPTPTTP-KEPASTTPKEPTPTTIKSAPTTTPKEP-APTTPKSAPTTTPKEP 262
Db      3560 lsekliipkkpklvrvprkeppakvpevkkiveekvrvpveeprrvpltkvpevlppkev 3619
QY      263 APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTTPKEPAPTTTPK-----PAPTTPKEPA 318
Db      3620 vp-----ekkvrvppakkeap-----pklvpeapkeevpevpevppkpkpevp 3665
QY      319 PTPPEPTPTTPTPKEPAPTTTPKEPAPTTTP-----KEPAPTPAPKAPPTTPKEPAPTTTPKE 371
Db      3666 tkvpevkaavpekkvpealiprkpespppevfespeespaaprk-kpevpvrivpevke 3724
QY      372 PAPTTKEPSPTTPKEPAPTTTPKSAPTTTPKEPAPTTTPKSAPTTTPKEPSPTTPKEPAPTTTP 431
Db      3725 vvpkekkyvaap--pkpdevrpvk-vpeapkeevpevpevpprk-ppkkvpevptk-----vp 3775
QY      432 KEPAPTPPKRAPPTTPKEPAPTTTPKEPAPTTTPKRAPATAKEPAPTTTPKETAPPTPKLT 491
Db      3776 evpkvavpekkvp-----ealppkpespppevfespeesvalaepeaevepeespaapqvlt 3831
QY      492 ----PTPEKLAPTT-----PEKAPTPPEELAPTTPEEPPT-TPPEEPAPTTPKAAPNT 542
Db      3832 vvpknvpekkapavakpkelppvkvpvpevpevpevpevpevppkvlvppk--peappakvpev 3869
QY      543 PKEPAP-----TTPEP-----APTTPKEPAPTTTPKETAPTTTPKGTAPTTTPKKEPAPTTTPK 592
Db      3890 pkevvppekvaavpkkevppakvpevkkpv-----leekpavp--vperaeespppevye 3942
QY      593 KPAPELAPTTTPKEPSTSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKKEPAPTTTPKGTAPTTTL 652
Db      3943 e--peelap-----eeelapeekvpvvaee-----eepevppavpeekklipkekvp-vl 3992
QY      653 KEPAPTTTPKRAPKELAPTTTPKGTSTSDKAPAPTTTPKETA-----PTTPKEPAPTTTPK 705
Db      3993 kkpbeapppkkepekev-----lekpklikrppppppapppkevdkekifqlkalipkkvpevp 4049
QY      706 KKP--APTTPETPTPTTSEVS--TPTTPKEP-----TTIHNSPDSSTPELSAETTPPA 754
Db      4050 qvpekelclpklvpggekkvkrkliprkpepkeevnlksvlrtirpeeeekve-----pkl 4105
QY      755 LENSPEKPGVPTTKTPAATKP-----EMTTAKDKTTERDLRTPTTTAAPKKTEATVT 810
Db      4106 le-kvkkpavp---eppppkpveevpvcvkrexllppltkvpeklpklpplpapepdk 4161
QY      811 TEKTESKITATITTYVTSTTTQDTTPFKITTLTKTTTLAKRVTTTKKTTTTEIMNK----- 866
Db      4162 pee-----evkltkppvepdpiaapvlpvvgkkaeak 4197
QY      867 --PEETAKPKDRATNSKATTPKPKOKPTKAPKKPTS--TKKPTTMPVRKPKTTTPP----- 918
Db      4198 apkeeaekpgrlpkypkktlpieaerklipsgsgkeppdeapfltyqlkavpklfveke 4257

```

```

QY 532 PTPKAAAPNT---PKPAPPTTPKPAPTTPKPAPTTPKGAAPTTPKPAAP 568
XX : : : : : : : : : : : : : : : : : :
XX 634 enrerlanekltsspaepengrtrpfanekltsspaephegertplanentlspaep 693
QY 589 -----TPKKRPAKELAPPTTKEPTSTSDKPAPTTKGTATPTTKEPAPT 642
XX : : : : : : : : : : : : : : : : : :
XX 694 tenrerlanektrfpaeprenrerlanentcspqplengrtrpfanekltspsaep 753
QY 643 -----TPKGTAPPTLKEPAPT-----TPKKPA-PKE-----LAP 670
XX : : : : : : : : : : : : : : : : : :
XX 754 engkrpfanekltsspaephegertplanentlsspaeprenrerlanekltqfpaep 813
QY 671 TTTKGPTSTSDKPAPTTKEPT-----APTTPKPAPTTPKPKAPPTPEPTPPPT 719
XX : : : : : : : : : : : : : : : : : :
XX 814 tenrerlanektrfpaeprenrerlanentlsspaephegertplanekltspsaep 873
QY 720 TSEVSTPTTKEPTTIKSPDE-----STPELSAEPPTK-----ALENSPKPEGV 764
XX : : : : : : : : : : : : : : : : : :
XX 874 engertpfnektspsaephegertplanentltps-reephegertlanekaltspak 932
QY 765 PT-----TKTPAATKP-----EMTTAKDKTTERDLTTPETTTAATPKMT 804
XX : : : : : : : : : : : : : : : : : :
XX 933 plehgetvnedltspsaepengertplanentltpstehgertlanekltspsaep 992
QY 805 KE-----TATTEKTTESKITAT-TOYTSSTTODTTPFXI-----TTL----- 842
XX : : : : : : : : : : : : : : : : : :
XX 993 tehgerlpsanekltspakplehgemtplanentlcpvkvplehgektlanekltsp 1052
QY 843 -----KTTTAPKVTYTT-----KTTTTEIMNKPEETAKPKDRATNSKATP 885
XX : : : : : : : : : : : : : : : : : :
XX 1053 eglehgaaktlanekltspakplehgerltspndkltssaaestehdratsanvlt 1112
QY 886 KPOKP-----TKAKKPTSTKKPTMPVKKP----- 912
XX : : : : : : : : : : : : : : : : : :
XX 1113 apaepihakrtllahakmtqvtleakstehektstle-Klttrpcklysektictg 1171
QY 913 KTPPTPKMTSTMPELNPTSRIAEAMLTQTTTRPNOT 948
XX : : : : : : : : : : : : : : : : : :
XX 1172 kntrvpektenlgnlttcltcltlakvksstenpekt 1207
Db

RESULT 12
AAW31852
ID AAW31852 standard; Protein: 763 AA.
XX
AC AAW31852:
XX
DT 27-APR-1998 (first entry)
XX
DE Mycobacterium tuberculosis 74 kDa protein.
XX
KW Tuberculosis; mycobacteria; infection; diagnosis;
KW antimycobacterial; antibiotic; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN W09741252-A2.
XX
PD 06-NOV-1997.
XX
PF 18-APR-1997; 97WO-EP01973.
XX
PR 29-APR-1996; 96DE-4017184.
XX
PI Espilita C, Honisch C, Moreno C, Singh M;
XX (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
XX
DR WPI: 1997-549750/50.
XX N-PSDB: AAT93610.
XX
PT New DNA and related proteins or RNA derived from M. tuberculosis -
used for diagnosis of mycobacterial infections, monitoring

```

```

PT vaccination and development of anti-mycobacterial agents
XX
XX Claim 5; Fig 13; 55pp; English.
PS
XX
XX This novel 74 kDa protein is encoded by an open reading frame of
CC a mycobacterium tuberculosis DNA fragment (see AAT93610) containing
CC polymorphic GC-rich sequences. Its amino acid sequence shows
CC a high proline content, but there is no homology to any known
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC proteins (see AAW31851-57) are claimed. These can be produced as
CC recombinant proteins, especially in bacterial, yeast, fungal or
CC higher eukaryote host cells, and used for diagnosing tuberculosis
CC and other mycobacterial infections in humans or animals. The
CC claimed proteins can also be used for epidemiological studies, for
CC monitoring vaccination, and for the development of vaccines and
CC anti-mycobacterial drugs.
XX
XX Sequence 763 AA:
SQ

Query Match 10.5%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. No. 5, 2e-35;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

QY 197 PTPKAEETTKGPAALTTPKEPTTPTPKPASTTPKKEPTPTTIKSAPTTPKPAPTTKSAP 256
XX : : : : : : : : : : : : : : : : : :
XX 3 pvp-----apalalalppappapaepekppfppap-----pappcwmvlvsap 46
QY 257 TTPKEPAPTTPKPAPTTPKPAPTTKKEPAPTTSAPTTPKPAPTTPKPAPTTPKE 316
XX : : : : : : : : : : : : : : : : : :
XX 47 pcp--pap-----papkpxskapfppvpappaparelalpp--pap-----peapre 90
Db
QY 317 PAPTTPKEPTPTT-----PKPAPATTKEPAPTTPK-EPAPTAPK-KPAPTTPKPAPTTPK 370
XX : : : : : : : : : : : : : : : : : :
XX 91 srpalppcoppvvlppdppeaapvpvpapnspfpfpfpapkftvpapvpv--pvpnsp 148
Db
QY 371 EP--APTTPKPSPTTPKKEPAPTTPKSAPTTPKPAATTPKSAPTTPKESPTTPKEPAP 428
XX : : : : : : : : : : : : : : : : : :
XX 149 ffpfpalalnpap-----kaplanspfpppapfpagf--ppaapvpvpapapkska 201
Db
QY 429 TTPKEPAPTTPKPAPTTPKPAPTTPKKEPAP-----TTPKKPAPTAPKPAPTTPKETAP 484
XX : : : : : : : : : : : : : : : : : :
XX 202 spptrppap-----pmpatpmeffpplppvppdpjsketppapppppfpfpavpplppvpl 256
Db
QY 485 TTPKKLTTPTPKEKAP-----TTPKEPAPTPEELAPPTPEEPTTPPEEPAPT-- 534
XX : : : : : : : : : : : : : : : : : :
XX 257 pvpnkltppap--appvavaavlvapocplpplnhnpappapavpvpplaplnshp 313
Db
QY 535 ---PKAAAPNTPKPAP----- 548
XX
XX 314 pappsapvpgvplaplpisgrfsvwksfclstfcccvcsgvlagalnpsrpsl 373
QY 549 --TTPKEPAPTTPKPAPTTPKETA---PTTPKGTA---PTTLKEPAPTTP-KKAPKE 598
XX : : : : : : : : : : : : : : : : : :
XX 374 tttpralpapfpplpplpplntavpplpplpplalappfpplpplspgvppap-- 431
Db
QY 599 LAPTTKEPTSTSDKPAPTTPKGTAPPTPKETA-----PTTPKEPAPTTPKGTATP 652
XX : : : : : : : : : : : : : : : : : :
XX 432 --plppgkpwltlpplapapepk--lvpylppgscppsekpnppapdeppeksspalpp 488
Db
QY 653 KEPAATTP---KKPAPKELAPTTTKGP--TSTTSOKPAPTTPEET---APTTPKPAPTTP 705
XX : : : : : : : : : : : : : : : : : :
XX 489 appapmpsaavrvppspfpappapapaaampalppappspalclcppfpappapnsp 548
QY 706 KKPAPTTPPEPTTPSEVSTPTTKEPTTIKSP---DESTPELSAEPPTKA--LENSP 759
XX : : : : : : : : : : : : : : : : : :
XX 549 --pappapfpklls--anppcpvppapnppappapppelpappdpfpfpvvaasp 604
Db
QY 760 KEPGVPTTKTPAATTPKETTAKDKTTERDLRTTPETTTAATPKATKATATTEKTESKI 819
XX : : : : : : : : : : : : : : : : : :
XX 605 --pappappappsalpfvnpa-----pfpcaapk-----srpal 637
Db
QY 820 TATTTQVSTTQDTTPPKITTLKTTTLAPKVTYTTTKTTTTEIMNKPEETAKPKDRATN 879

```







PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140655.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142154.  
 PR 06-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142350.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 03-AUG-1999; 99US-0146389.  
 PR 04-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149375.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150565.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match

14.4%; Score 981; DB 21; Length 763;

(e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences CC given in the exemplification of the present invention.

Sequence 5179 AA:

Query Match 17.2%; Score 1171; DB 22; Length 5179;  
Best Local Similarity 32.8%; Pred. No. 2.3e-61;  
Matches 364; Conservative 57; Mismatches 461; Indels 228; Gaps 41;

38 ERGRCDCAQ---CKRYDK-----CCPDYE--SFCAEVKDKKKKKKKPKPPPV 84  
1350 qhgqgqvcdsvagficknedqfgnpgfgydklrvcwcpmd-----kcltppspc 1403  
85 VDEAGSGLDNGDFKVTTPDTSTQHNKVSSTPKITTAAPINRPSLPNSDSKE--TSL 142  
1404 ttp-----spptltt-----tlpdtppspptlttttppdtppspptlt 1445  
143 TVNKETTVETK--ETTTTNKQSTDEKETSASKEQSIKTSANDLADPTSKVLAKPPPK 200  
1446 ttpdtppspstlttppdtppspptlttppdtppspptltt-----pspp 1487  
201 AETTTKGALTTPEKP--TPETPKPEASTTPKEPTTTKSAPTTPKEBAPTTPKSAPT 258  
1488 ttttppdtppspmtptp--pactltlpt--tppsptlttppdtppsppt 1544  
259 PKRPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPT 318  
1545 tptpdtppstlttppdtppspptlttppdtppspptltt-----tlpdtppsp 1598  
319 PTPPKPEPTTPKEBAPTTPKEBAPTTPKEB-----APPAKKAPPTTPKEBAPTTPKE 371  
1599 ttpspptlttppdtppspptlttppdtppstltlppdtppspptlttppdtpps 1658  
372 PATTTKEBPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAP--TPKEBPTTPKEBAPT 428  
1659 ptttppspptlttppdtppspstlttppspptlttmtppstlttppspstlttpps 1717  
429 TTPKEBAPTTPKKRPAPTTPKEBAPTTPKEBAPTTPKKRPAPTTPKEBAPTTP--KET 483  
1718 tttppspptmtppsppttppspptlttmttppstlttppstlttppstltt 1777  
484 PTPP-----KRLTP----- 492  
1778 ptpcpvplcmwtgldsqkfnhpggdtelldgcvgpwaanlsctraunypvbjgld 1837  
493 -----TTPKLAPTTPKEBAPT 510  
1838 qtvvcadvagficknedqkpgvlpmafclynelnvqcecvltumtltttenpctpc 1897  
511 PEELAPTPPEBPTT-TPPEBAPTTPKAABNTPKEBAPTTPKEP-----APTTPKEBAPT 564  
1898 tpttlttvtprpcpctgctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 1957  
565 T-TPKEBAPTTPKGTAPTTLKEBAPTTPKKAPKELAPTTTKEBPTSTSDKPAP--TPPKG 622  
1958 tptgtcgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2017  
623 TAPTPKEBAPTTPKEBAPTTPKGTAPTTLKEBAPTTPKKAPKELAPTTTGTSTTSD 682  
2018 tttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2076  
683 KPAP--TPPKELAPTTTPKEBAPTTPKKAPKELAPTTTPPEBPTSTSVSPPTTKBPTTHKSPDE 741  
2077 tptpdtppstlttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtct 2132  
742 STELSAPTPKALENSPKP-----GVPT--TTPATKPEMTTAKDXTTERDL 790  
2133 tttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2192  
791 R---TTP--ETTTAAPKMT---KETATTEKTESKITAATTQVTSITTTQDTTPPKITTL 842

Db 2193 qtpdtppdtlttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtct 2248  
Qy 843 KTTTLAPKVTYT--KTTTTEIEMKKEPETAKPDRAINSKATTPPKPKAP--KRPST 900  
Db 2249 -ttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 2307  
Qy 901 KKPETMPVRKPKTTPPKMTSTMPKELNP--TSRIAEMLQTTTP--PNQTPNSKLVEY 956  
Db 2308 qtpdtppdtlttvtprpcpctgtctgtctgtctgtctgtctgtctgtctgtctgtct 2367  
Qy 957 NPKSEDAGAGETPPHMLLRPHVMEVTP 986  
Db 2368 ttptrpctg---tqtp--tltppdtlttvtpr 2392

RESULT 9  
AAG38942  
ID AAG38942 standard; Protein: 763 AA.  
XX  
AC AAG38942;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135533.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 395..396
XX FT Misc-difference 444..446
XX FT Misc-difference 444..446
XX FT /note="unspecified amino acids"
XX FT /note="unspecified amino acids"
XX PN WC9523861-A1.
XX PD 08-SEP-1995.
XX PE 06-MAR-1995; 95WO-CN00015.
XX PR 04-MAR-1994; 94CN-0112066.
XX PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX PI Gu X, Han Z, Shen Q;
XX DR MPI; 1995-320576/41.
XX DR N-PSDB; AAT04546.
XX PT New haematopoietic cell growth factor - used for treating
XX PT thrombocytopenia and hematocytopenia
XX PS Example: Page 23; 36pp; Chinese.
XX XX
XX CC This sequence represents the human megakaryocytopoietin (MPO) protein.
XX CC This sequence was purified using a carrier which can couple wheat germ
XX CC agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX CC AAR80039 and AAR80040) were used to produce the amplification primers
XX CC shown in AAT04544 and AAT04545. The fragments amplified by these primers
XX CC can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX CC The MPO cDNA can then be inserted into a plasmid which is used to
XX CC transform cells to produce MPO. The MPO sequence is capable of promoting
XX CC colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX CC and stimulating the proliferation of multipotential stem cells. The
XX CC factor may be used for treating thrombocytopenia and hematocytopenia.
XX CC The purification method can be used to isolate MPO from human urine or
XX CC serum of patients with aplastic anaemia, and from animal blood or urine
XX CC by radiation exposing the animals to induce aplastic anaemia.
XX SQ Sequence 452 AA;

```

```

Query Match 24.3%; Score 1654; DB 16; Length 452;
Best Local Similarity 72.2%; Pred. No. 1.9e-91;
Matches 328; Conservative 3; Mismatches 21; Indels 102; Gaps 3;

```

```

QY 1 MAWTLPIYLLLSVFVIOOVSSO----- 25
DB 1 MAWTLPIYLLLSVFVIOOVSSO----- 25
QY 26 -----ELISCKRPFSEFERGECDDCAQCKKYDKCCPDYESFCA----- 64
DB 61 krvtlaelsckgrcfesfergredcdagckkydkccpdyesfaevhnmpssskap 120
QY 65 -----EYKDNKKKNTKKKKPPKPPVVDAG 89
DB 121 ppsgaagtlstcktrspkppnkkkkkviseellevkdkhknrtkkktpkppvdeag 180
QY 90 SGLDNGDFKVTPTSTTOHNVKSTSPKITAKPINRPSLPNSDPSKSTLVNKKET 149
DB 181 sglngdfkvtptsttqhnkvstspkltakpnrpslpnsdpskstsitvnhelt 240
QY 150 VETKETTTKOTSTDEKKTTSKKEQSTLEKTSANDLAPTSVLAKPPTPAKTTTGGPA 209
DB 241 vetketttkqstcdqkeltakelqstakdlaptskvlakpkaetttkqpa 300
QY 210 LTPPKPEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTPKEPAPTTKE 269

```

```

DB 301 ltpkpepttpkpeasttpkpepttllksaptlpkpeapttksaptlpkpepttke 360
QY 270 PAPTTPKEPAPTTPKEPAPTTKSAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPTT 329
DB 361 papttppkpeaptttkpeapttkshpdlprscxxxxcqp-----tpephppl 409
QY 330 PREPAPTTPKEPAPTTPKEPAPTTPKPAPTTPKE 363
DB 410 prslhptkpeaptttkpeaptpkpaaplpplle 443

RESULT 8
AAM24516
ID AAM24516 standard; Protein: 5179 AA.
AC AAM24516;
XX AA
XX 12-OCT-2001 (first entry)
DE C899P predicted amino acid sequence.
XX DE
XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
XX KW immunogenic; gene therapy; vaccine; colonic cancer.
XX OS Homo sapiens.
XX PN WC200149716-A2.
XX PD 12-JUL-2001.
XX PE 29-DEC-2000; 2000MO-US35596.
XX PR 30-DEC-1999; 99US-0476296.
XX PR 10-JAN-2000; 2000US-0480321.
XX PR 15-FEB-2000; 2000US-0504629.
XX PR 06-MAR-2000; 2000US-0519444.
XX PR 19-MAY-2000; 2000US-0575251.
XX PR 29-JUN-2000; 2000US-0609448.
XX PR 28-AUG-2000; 2000US-0649811.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;
XX PI King GE, Wang T, Jiang Y;
XX DR WPI; 2001-441847/47.
XX PS Claim 2; Page 446-462; 472pp; English.
XX XX
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP)
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying mutations or deletions in a patient's genome that affect the
XX CC activity of TCAPs by expressing inactive proteins or to supplement the
XX CC patient's own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs
XX CC and in assays to identify modulators of TCAP expression and activity.
XX CC Anti-(I) antibodies and antagonists may also be used to down regulate
XX CC TCAP expression and activity. The anti-(II) antibodies may also be used
XX CC as diagnostic agents for detecting the presence of TCAPs in samples

```



```

Db 601 aptapkeapltckepaltcpkkltpctpeklapltcpkeapltcpelapltcpcept 660
Qy 527 PEEBAPPTPAAAPNPTKEBAPPTPKBAPPTPKBAPPTPKATTTKGAATTTLKEP 586
Db 661 peepapltcpkaapnrcpkeapltcpkeapltcpkeapltcpkeltapltckaplt 720
Qy 587 APTTPKPAKELAPTTTKEPTSTSDKAPPTPKGTATPTPKBAPPTPKBAPTPG 646
Db 721 apttpkpkpkeapltckepstsdckapltcpkeltapltcpkeapltcpkeapltcp 780
Qy 647 TAPTTLKEBAPPTPKPAKELAPTTTCKPTSTSDKAPPTPKBAPTTPKBAPTPPK 706
Db 781 tapttlkeapltcpkpkpkeapltckpeltscdckpaptcpkeapltcpkeapltcpk 840
Qy 707 KPAPTTPETPTTSEVPTTPTTKEPTTHKGPDESTPELSAEPPTKALENSPKBEPVT 766
Db 841 kpaaptcpcepprttsevspttckepcttknspdeetcpelaseaptpkalenspkbpvt 900
Qy 767 TKTPAATKEBMTTAAKDKTTERDLRTPTPTTAAAPKMKETATTTETKESKITAATTQV 826
Db 901 ktpraekpemtackdkltetrdlttptetlaapkmcketalettekstetatttq 960
Qy 827 TSTTQDTPTPFITTLTKTTTLAPKVTYTKTTTTEINMKPEETAKKPDRAITNSKATTPK 886
Db 961 ttttqdtptpfittlttktttlapkvctckklttletlmkpeetakpdratnskattpk 1020
Qy 887 POKPTPAKPKPTSTKPKPTMPRVKPKTTPPKMTSTMPELNPSRTAEMLOTTPRPN 946
Db 1021 pdkpkpapkpkstckkptkmpvrkpkcttpkmtscmpelnpsrtiaeemlottpn 1080
Qy 947 QTPNKLVEVNPKEBAGGEPHMLRHPVPEVTPKMDVLPVPCNGIITNPLS 1006
Db 1081 qtpnklvevnpkesaggeepcmllrhpvimevtpmdylpvpcngiitnplls 1140
Qy 1007 DETNICKNGKPVGDLTTLNGLVAFRGHYFMMLSPFSPSPARRTTEWGISPIDVET 1066
Db 1141 detnckngkpvdglttllnnglvafrghyfmmlspfsparrlttewgispidvft 1200
Qy 1067 RCNCGKTEFFKDSQYWFMTNDIDAGYPRPIKFGGGLGQIVAAIATAKYNMPESVY 1126
Db 1201 rcnckgkttfkdsqywfntndidagypkplfkfggltgqivaalstakynmpesvy 1260
Qy 1127 FFKRGGSIQOYIYKQEPKQCPGRRPALNYPVGE 1161
Db 1261 ffkrggsiqylykqepvkcpqgrrpalnypvge 1295

```

RESULT 5  
AAB29778 standard; Protein; 902 AA.

AC AAB29778;

DT 28-FEB-2001 (first entry)

DE Human MSF-derived tribonectin.

KM Human tribonectin; MSF: megakaryocyte stimulating factor;  
KM alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KM osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KM friction coefficient reduction; gene therapy; antiarthritic;  
KM osteopathic.

OS Homo sapiens.

PN M0200064930-A2.

PD 02-NOV-2000.

PF 24-APR-2000; 2000MO-US10953.

PR 23-APR-1999; 9905-0298970.

```

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
PA Jay GD:
PI WPI: 2001-024673/03.
XX Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety
XX
PS Disclosure; Fig 1; 47pp; English.
XX
XX The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76
CC repeats of a motif having at least 50% identity to the sequence KEPAPTT
CC (AAB29774). The invention also relates to a nucleic acid encoding a
CC human MSF-derived tribonectin; a biocompatible composition comprising a
CC human tribonectin for inhibiting tissue adhesion formation; and a method
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
CC treatment of osteoarthritis, where they may be used for lubricating
CC mammalian joints, such as articulating joints of humans, dogs or horses.
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
CC useful for inhibiting adhesion between two surfaces such as the injured
CC tissues of a mammal, where the injury is caused by a surgical insertion
CC of trauma, or an artificial device e.g., an orthopaedic implant. In
CC particular, one of the surfaces is pericardial tissue. DNA encoding a
CC tribonectin may be used in gene therapy. The present sequence represents
CC a substantial portion of a human MSF-derived tribonectin.
XX
SQ Sequence 902 AA:

```

Query Match 51.1%; Score 3484; DB 22; Length 902;  
Best Local Similarity 74.6%; Pred. No. 1.3e-200;  
Matches 794; Conservative 28; Mismatches 80; Indels 162; Gaps 66;

```

Qy 66 VKDNKKNRTKKRPPKPVVDEAGSLDNGDFKVTTPDSTOHKNVSPKITTAKPTN 125
Db 1 vkdnknrkkkpkpvpvdeagsgldngdfkvttpdstohknvksptiltakpn 60
Qy 126 PRPSLPNSDTSKETSIVNKKETVETKTTTNKQSTDGKEKTTSAKETOSIEKTSK 185
Db 61 prpslpnsdtsketsilvnkeltvetketttnkqstdgkelttsakegsiektak 120
Qy 186 DLAPTSKVLAKPPTKAEETTKGPALTTTPKEPTTPPKBPASTTPKEPTTTKSAPTTPK 245
Db 121 dlaptskvllakpptaetttkgpaltttpkepttpkbpaasttpkeptttksapttpk 172
Qy 246 EPAPTTTSAPTTPKEBAPTTTKEBAPTTTPKBPATTTKEBAPTTTKSAPTTPKBAPTT 305
Db 173 epaptttsapttpkebaptttkebaptttpkbpatttkebaptttksapttpkbaptt 228
Qy 306 PKKPAPTPPKBAPTPPKBAPTTTPPKBAPTTTPKEBAPTTTPKEBAPTTAKKAPTPPKBPA 365
Db 229 -keapltc-keapltc-keapltc-keapltc-keapltc-keapltc-keapltc-kepa 281
Qy 366 PTPPKBAPTTTPKEBAPTTTPPKBAPTTTKSAPTTTKBAPTTTKSAPTTKEBAPTTTKE 425
Db 282 pttpkbaptttkebaptttkebaptttksaptttkbaptttksapttkebaptttke 335
Qy 426 PAPTTPKEBAPTTTPKKAAPTTPKEBAPTTTPKEBAPTTTKKAPAPAKBEPATTPKETAPT 485
Db 336 papttkebaptttppkkaapttpkebaptttkebaptttkkapapakbepattpketapt 388
Qy 486 TPKKLTPTTPPEKLAPTTPPEKPAPTTPEELAPTTPEEPPTTPPEBAPATTTKAAAPNTPKE 545
Db 389 t-kcapltckep-apttke-papltc-keapltc-keapltc-keapltc-keapltc-ke 440

```

|||||  
Db 721 aqtcpxkpkapkeaplttckptsdsdpapltgqtpctpckepapltcpkcpapltcpk 780  
QY 647 TAPTLKEBPAPTPPKPAKELAPTTTGGPTSTSDKAPAPTPPKTAATTPKEBPAPTPK 706  
Db 781 taptlckepapltcpkpkapkeapltckptsdsdpapltcpkcpapltcpkcpapltcpk 840  
QY 707 KAPAPTPPEPPTTSEVSPPTTKEPPTTHKSPDESTPELSAEPKALENSPEKPGVPT 766  
Db 841 kpaaptpceppttsevspttckepctlnkspdesctpelasepckalenspkpgvpt 900  
QY 767 TETPAATKPEMTTAKDKTTERDLRTPEPTTAAPKMTKEATATTEKTTESKITATTOY 826  
Db 901 tktpatckpemtltakdkctterdlrttpepttaapkmktetatektskitatttqv 960  
QY 827 TSTTTODTTPFKITTLKTTTLAPKVTMTTKTTTTEIMKPEPTAKPKDRATNSKATPCK 886  
Db 961 tstitdttptfiktltklttlapkvttckltteltimkpeetackpkdratnskatpck 1020  
QY 887 POKPTAKPKPTSTKPKPTMPRVKPKPTTPPKMTSTMPELNPTSRIAEAMLQTTTPRN 946  
Db 1021 pqkptapkkptstckpkmpvrktpcktpckmtstmpelnpstriaeamlqtttrpn 1080  
QY 947 QTPNSKLIVEVNPKSEDAAGBETPHMLLRPHVFEVYPPDMOYLPRVNOGIIINPMLS 1006  
Db 1081 qtpnslkivevnpksedagagsetphmlrrphvfmevtpdmoylprvnpqgillnplms 1140  
QY 1007 DETNINCGKVPDGLTTLRNQTLVAFRGHYFWMLSPFSPSPARITIEWGIPSPIDTVEF 1066  
Db 1141 detnlcnqkvpdglcttlrnqtlvafgrhyfwmfslpsfssparitiewvgispdttvtf 1200  
QY 1067 RCNCEGTFPEFKDSQVWRFTNDIKDAGYPRIFKFGGLTGQIVAAALSTAKYNNMESY 1126  
Db 1201 rcncegtrffkdsqyvrftndikdaqypkrfkfglqgivaalstakynmesy 1260  
QY 1127 PFKRGGSIQOYIKQEPVOCQGRPALANYPVGEVORRRRFEERAIIGSQTHTIRIQY 1186  
Db 1261 ffkrggsiqgyikqepvqcpgrpalanypvgemqvrtrrrfeteraigsqhtirigy 1320  
QY 1187 SPARLAAYQDKGVLHNEKVSILMGRLPNVYTSALSPNIRKPDGYDYAASQDYNYNDY 1246  
Db 1221 sparlalqdkgvlhnekvslmrglpnvvtalslpnirkpdgydyatskdqynndv 1380  
QY 1247 PSRTARAITTSSGOTLSKVMYNCP 1270  
Db 1381 psrtaratlrrsgqlskvmyncp 1404  
RESULT 4  
ID AA024322 standard; Protein: 1299 AA.  
XX AC AA024322:  
XX 12-OCT-2001 (first entry)  
XX DE Human EST encoded protein. SEQ ID NO: 1847.  
XX XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
OS Homo sapiens.  
XX OS  
XX PN WO200154477-A2.  
XX PD 02-AUG-2001.  
XX PF 25-JAN-2001; 2001WO-US02687.  
XX PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX WPI: 2001-476164/51.  
DR N-PSDB: AAH98981.  
XX  
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
PS antibodies and research use -  
PS Claim 20; Page 1198-1201; 1275pp; English.  
XX  
XX CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
XX SO Sequence 1299 AA;  
Query Match 90.3%; Score 6151; DB 22; Length 1299;  
Best Local Similarity 89.7%; Pred. No. 0; Mismatches 0; Indels 134; Gaps 2;  
Matches 1161; Conservative  
QY 1 MAMKTPYLILLLSVFIQVSSO----- 25  
Db 1 mamktpllyllllsvfivqvsqdlscagrcgegyrdatcncdynqihmecpbf 60  
QY 26 -----ELSCGRCFSEFGRGRCDCDAQCKRYDCCPDYSEFCAE----- 65  
Db 61 kryctaelockgrcfesfegrccdaqckrydcpdysefcaevhnptspsskkap 120  
QY 66 ----- 65  
Db 121 ppsgaqtlkstktrkpkpnnkkkkkvieseeleehsvsengeassssssssstlw 180  
QY 66 -----VKDNKKKRTKKKPPRPVYVDEAGSGGNDKDFVYTPDST 106  
Db 181 kiksksnaaanrelqkkklykdknrtkkkpkpvpvdeagsgldngdtkvtpdst 240  
QY 107 TOHNKYSTPKITTAPINRPSLPNSDTSKETSLVKNKETTVEKETTNNKQTSIDG 166  
Db 241 tqhnkystpskittakpindrpslppnsdstsketsltnkettvetkettnkqstidg 300  
QY 167 KEKTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPAULTTPKEPTPTTKEPAS 226  
Db 301 kektsaketgsiektsakdlaptskvlakpdpkaetltkypalttppkeptpttkepas 360  
QY 227 TTPKEPTPTTTSKAPTPPKKEPAPTTTTSAPTPPKKEPAPTTTKEPAPTTTKEPAP 286  
Db 361 ttpkeptptttskaptppkkepaptttsaptppkkepaptttskeptpttkepapttkep 420  
QY 287 APPTTKSAPTPPKPAPTPPKKAPPAITTPPKAPPTTTPPKKEPAPTTTKEPAPTTTKEP 346  
Db 421 appttksaptppkpaptppkkappaaitppkapppttppkkepaptttskepapttkep 480  
QY 347 EPAPTAAPKAPPTPKKEPAPTPPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 406  
Db 481 epaptapkpapttpkkepapttpkkepapttkepapttkepapttkepapttkepapt 540  
QY 407 TTKSAPTPPKESPPTTKEPAPTPPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 466  
Db 541 tksaptpkpspttkepapttpkkepapttkepapttkepapttkepapttkepapttkep 600  
QY 467 APAPKKEBPAPTPPKETAPTPPKLLPTTPPEKLAFTTEBKAPATTPEELAPTTPEPTPT 526





OY	1127	FFKRGSGTIOQYIYKOEYVQKCPGRPALNIPYGGMTQYRRRRFERAIGPSTHTIRIOY	1186
Db	1261	ffkrgsgtqyiykpeypqkcpgrpalnipyggmtqyrrrrferaiqpsqthiriy	1320
OY	1187	SPARLAQDDKCVLHNEVVSILMREGLPMVWVSALSLPNIKRKPDGXYAFSKDQYTNIDV	1246
Db	1321	sparlaydkcvlhnvsvsilmrslpmvwsalslpnirkpdgydyafskdytynidv	1380
OY	1247	PSRTARAITTRSGQTLISKWYNCP	1270
Db	1381	psrtaraittrsgqlskwyncp	1404
RESULT	2		
ID	AAB60568		
XX	AAB60568	standard; Protein; 1404 AA.	
XX			
XX	AAB60568;		
XX			
XX	27-APR-2001	(first entry)	
XX			
XX	DE		
XX		Human megakaryocyte stimulating factor (MSF, CACP).	
XX			
XX	Human, CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;		
XX	MSF megakaryocyte stimulating factor; synovial lubricant;		
XX	chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;		
XX	antiarthritic.		
XX			
XX	Homo sapiens.		
XX			
XX	WO200107068-A1.		
XX			
XX	01-FEB-2001.		
XX			
XX	21-JUL-2000;	2000OWO-US20002.	
XX			
XX	23-JUL-1999;	99US-0145328.	
XX	19-JUL-2000;	2000US-0145328.	
XX			
XX	(UYCA-) UNITV CASE WESTERN RESERVE.		
XX			
XX	Warman ML;		
XX			
XX	WPI; 2001-182721/18.		
XX			
XX		New composition comprising the camptodactyly-arthropathy-coxa	
XX		vara-pericarditis protein in combination with an anesthetic, useful for	
XX		treating osteoarthritis, or as lubricants of tissue and joints	
XX			
XX	Example 1; Page -: 34pp; English.		
XX			
XX		The invention relates to a method of treating osteoarthritis via the	
XX		administration of a composition comprising the camptodactyly-arthropathy-	
XX		coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.	
XX		The composition may further comprise a local anaesthetic. The composition	
XX		of the invention may be administered via intra-articular or intravenous	
XX		injection. The human CACP protein is identified in the invention as	
XX		being megakaryocyte stimulating factor (MSF). The gene encoding	
XX		CACP protein (MSF) is located on chromosome 1q25-31, and mutations in	
XX		this gene are responsible for the heritable disorder camptodactyly-	
XX		arthropathy-coxa vara-pericarditis, in which patients have synovial	
XX		hyperplasia without evidence of inflammation. CACP protein (MSF)	
XX		acts as a synovium lubricant, and can be used to lubricate tissue and	
XX		joints in the treatment of osteoarthritis. The composition may be	
XX		applied to reduce the symptoms of osteoarthritis (e.g., joint pain,	
XX		loss of range of movement or joint damage). The present sequence	
XX		represents human megakaryocyte stimulating factor (MSF, CACP protein).	
XX		Note: This sequence is not given in its entirety in figure 4 of the	
XX		specification, although a GenBank accession number was given. This	
XX		sequence was therefore obtained from GenBank (U70316).	
XX			
XX			
XX	Sequence	1404 AA;	
SQ			

Query Match	98.7%	Score 6127	DB 22	Length 1404	
Best Local Similarity	90.5%	Pred. No. 0			
Matches 1270	Conservative 0	Mismatches 0	Indels 134	Gaps 2	
QY	1	MAMKLPITYLILLLSVFVIOOVSSO-----			25
DB	1				60
QY	26	-----ELSCNGRCSESEFERGREGDCDAOCKYKDKCCPDYESFCAB-----			65
DB	61	kvrcfaelsckgrcfesfergreccdaqckkydkccpdyesfcaeavhnlpspskkap			120
QY	66	-----			65
DB	121	ppsgasqlstktsrpkppnkkkkkkvieesieiehvnsengessssssssstlw			180
QY	66	-----VKDNKKNRTKKKPKPKPPVYVDEAGSLDNGDFKVTYPTDTST			106
DB	181	kllksksnasanelqkllkvdkdknrrtkkkppkpvvdaagsglidnglflkvtlpdst			240
QY	107	TOHNKVSTSPKITTAKPINPRLPSLPNSDTSKRTSLTVNKETVLEKRTTTHKKQSTDG			166
DB	241	tqhnkvsrpklttckpimprlpnsdstksclstlvnkvettvetcttnkqstgd			300
QY	167	KEKTTSAKENGOSTIEKTSAKLAPTSVYLAKPKPAETTTGKAPLITPKEDPTTPKBPAS			226
DB	301	kektsaketslektsakdlaptskvlakpikpaeelttqpalitupkeptitupkepas			360
QY	227	TPPKPEPTPTTKSAPPTPKPEBAPTTTKSAPTTTKKEAPTTTKKEBAPTTTKKEB			286
DB	361	tlupkeptitkapttkpeaplttkksaplttkpeaplttkpeaplttkpeaplttkpe			420
QY	287	APTTSAPPTPKBPAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK			346
DB	421	aplttksaplttkpeaplttkkbpaplttkpeaplttkpepttkpepttkpeaplttkpeaplttk			480
QY	347	EPAPTPAPKKBPAPTTPKBPAPTTPKBPAPTTKEBSPPTPKBPAPTTTKSAPTTTKBPAPT			406
DB	481	epaptpapkkbpaplttkpeaplttkpeaplttkespttkpeaplttkksaplttkpeaplt			540
QY	407	TTTSAPPTPKBESPPTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP			466
DB	541	tlksaplttkpeaplttkpeaplttkpeaplttkpeaplttkbpaplttkpeaplttkpeaplttkkp			600
QY	467	APTAPKEBAPTTPKETAPTTPKLPTTPEKLABTPEKAPATTPEELAPTTPEEBPTT			526
DB	601	aptpapkeaplttkpeaplttkltpitpeklaplttkpeaplttkpeaplttkpeaplttkpeaplttk			660
QY	527	PEEPAPPTPKAAAANTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT			586
DB	661	peepaplttkaaaaplttkpeaplttkpeaplttkpeaplttkpeaplttkcaplttkpalttklkep			720
QY	587	APTTPKKBPAPKELABTPTTKKEPTSTTSDBKPADTTPKGTAPTTPKBPAPTTPKBPAPTTPKG			646
DB	721	aptpkpapkelaplttkkeptsttsdskpalttkgtaplttkpeaplttkpeaplttkp			780
QY	647	TAPPTLKEBAPTTPKKAPKELATTTKGPSTSDKAPTTPKETATTPKEBAPTTPK			706
DB	781	taplttkpeaplttkkbpakelaplttkgprstsdskpalttkpeaplttkpeaplttkp			840
QY	707	KPAPTPEPTTPPTSEVSTPTTKEBPTTHKSPDESTPELSAETPKALENSPKREGVPT			766
DB	841	kpapttkpepttksevspttkkepttkhspdesptelsaepkpalenspkregypt			900
QY	767	TKTPAATPKEMTTAKDKTTERDLKTPETTYAAPKMTKEATATTEKTESTTATTTTOY			826
DB	901	tktpaatkpmnttakaktterdlrttpebtlaapkmketaattekteslftatltqv			960
QY	827	TSTTTQDTPPKKITLKTATTLAPKVTTKKITTTEINMKPEETAKPKDRAVNSKATTPK			886
DB	961	tstttqdtppklttklktlatpvttkklttiteimkpeetakpkdrravnskattpk			1020

```

FT      Region      1266..1331
FT      /label= Exon_X
FT      Region      1331..1373
FT      /label= Exon_XI
FT      Region      1373..1404
FT      /label= Exon_XII
XX      WO9213075-A.
XX      06-AUG-1992.
XX
XX      17-JAN-1992; 92WO-US00433.
XX
XX      18-JAN-1991; 91US-0643502.
XX      10-SEP-1991; 91US-0757022.
XX      (GEMV ) GENETICS INST INC.
XX
XX      Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;
XX      WPI; 1992-284660/34.
XX      N-PSDB; AAQ27223.
XX
XX      New human mega-karyocyte stimulating factors - for treating
XX      immune deficiencies, cancer, exposure to radiation or drugs,
XX      bacterial and viral infections, etc.
XX
XX      Claim 1, 2 and 3; Fig 1; 87pp; English.
XX
XX      The sequence given is a full length translation from the megakaryocyte
XX      stimulating factor (MSF) precursor. The sequence covered by exons II,
XX      III and IV encodes megakaryocyte stimulating factor (MSF). This
XX      sequence is modified by the addition of an N-terminal sequence encoding
XX      a secretory leader, an initiating methionine proceeding exon II and a
XX      terminating codon following exon IV. The cDNA sequence given contains
XX      sequences derived from human megakaryocyte colony stimulating factor
XX      (meg-CSF). Exon I contains the initiating methionine, and encodes a
XX      classical mammalian protein secretion signal sequence. The sequence
XX      encoding the original meg-CSF includes exons II-IV and is thought to
XX      terminate in the region between amino acid residues 134 - 147. The
XX      yield a family of mRNAs each encoding a different MSF protein. Exons
XX      V and VI are thought to be related to the activity of the factor and
XX      are also implicated in the stability, folding and processing of the
XX      molecule. These exons are also thought to play a role in the observed
XX      synergy of MSF with other cytokines. Exons V - XII are believed to be
XX      implicated in the processing or folding of the appropriate structure of
XX      the resulting factor, ie. one or more of these exons may contain
XX      sequences which direct proteolytic cleavage, adhesion, organisation of
XX      the cellular matrix or extracellular matrix processing. Both naturally
XX      occurring and non-naturally occurring MSF's may be characterised by
XX      various combinations of alternatively spliced exons from this sequence,
XX      members of the MSF family.
XX
XX      Sequence 1404 AA:
XX
Query Match      98.7%; Score 6727; DB 13; Length 1404;
Best Local Similarity 90.5%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

```

```

QY      66 -----VKDNKKNRKKRPKPPVVDAGSGLDNGDFVTTPTST 106
DB      181 k1skksnaaneqlqk1kvdknknrckkrcprrpvvdaagsgldngdfkvltptstc 240
QY      107 TOHNRVSTSPKITTAKPPIPNRPSLPPNSDTSKETSLSVYANKETTYETKETTNNKQTSYD 166
DB      241 tqhnkvstspkittcakrnprrpslppnsdtskselslvtvnetkettltnkqtsd 300
QY      167 KEKTSKAKETOSIEKTSKADIAPTSKVLAKEPPEKAEETTKGPAITTPKEPTTPKBPAS 226
DB      301 kektsakelqtsiektsakdiaptskvlakeppekaeettkgpaittpkepttpkbpas 360
QY      227 TTPKEPTTPKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 286
DB      361 ttpkepttpksaptpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 420
QY      287 APTTTSAPTPPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 346
DB      421 apttksaptpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpk 480
QY      347 EPAPTAKKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 406
DB      481 epaptakkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 540
QY      407 TTSAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 466
DB      541 ttsaptpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbp 600
QY      467 APTAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 526
DB      601 aptapkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpkbpatt 660
QY      527 PEPAPATPKAAAPNTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBP 586
DB      661 peepapatpkaaapntpkbpattpkbpattpkbpattpkbpattpkbpattpkbpattpk 720
QY      587 APTPKKPKAPKELAPPTTKBPSTSDKRAPTPPKGATTAETTKBPATTPKBPATTPKBP 646
DB      721 aptpkpkpkelappttkbpstsdkraptppkgatttaettkbpattpkbpattpkbpatt 780
QY      647 TAPTLLKEBPATTPPKBPAPKELAPPTTKGPTSTSDKRAPTPPKGATTAETTKBPATTPK 706
DB      781 taptllkebpattppkbpapkelappttkgptstsdkraptppkgatttaettkbpattpk 840
QY      707 KPAPTPEPPTSEVSTPTTKKPTTIHKSPEDESTPELSAEPKALENSPEKPEGYPT 766
DB      841 kpaptpepptssevstpttkkpttihkspestpeelsaepkalenspekpegypt 900
QY      767 TTPPATRPEMTTAKDKTTERDLRTTPEPTTAAPKMTKETATTEKTTESKITATTTQV 826
DB      901 ttppatrpemttakdktterdlrttpepttaapkmtketatttektteskittatctqv 960
QY      827 TSTTQDTPPKKITTLTKTTTTLAPKVYTTKKTITTEIMNKPEETAKPKDRATNSKATTPK 886
DB      961 tsttqdtpkkitltlkttttlapkvyttkkittteimnkpeetakpkdratnskatctpk 1020
QY      887 POKPTAKKRPSTSTKPKTPMPVRKPKTTPPKMTSTMPRLNTSTSLAEMLOTTTRPN 946
DB      1021 pkpptakkrpststkpktpmpvrkpktpprkmtstmprlntstslaeamlootttrpn 1080
QY      947 QTPNSKLVEVNPKSEDAGAGETPPHMLLRPHVMPREVTPTDMDVLPVNPQGIINPMLS 1006
DB      1081 qtpnsklvevnpskedagagetpphmlrrphvmprevtptdmdvlpvnpqgiiinpmls 1140
QY      1007 DETNICKGKRPVDGTTJLRNGTLVAFRGHYPMULSPSPSPARIRTWGICSPIDTVFT 1066
DB      1141 detnickgkrpvdgttljrngtlvafrghyfmlspspsparirtwgicspidtvft 1200
QY      1067 RCNCEGKTFFFKDSQWRFTNDIKDAGYPRPIFKGSGGLGQIVAAALSTAKYKNMPESTY 1126
DB      1201 rcncegkttffkdsqwrftndikdagypripfksgsglglqivaaalstakymnpeesty 1260

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:24:31 ; Search time 138.84 Seconds  
(without alignments)  
677.565 Million cell updates/sec

Title:	AA3
perfect score:	6814
Sequence:	1 MAKKTLPIYLLLSVFVIO.....ARAITRRSGQLSKWYNCP 1270

Scoring table: BLOSUM62  
Gapop 10.0 / Gapext 0.5

```
Searched:      522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

```
1: A_Geneset_1101.*
2: /S1D58/gcgdata/genseq/genseq/AAL1980.DAT.*
3: /S1D58/gcgdata/genseq/genseq/AAL1981.DAT.*
4: /S1D58/gcgdata/genseq/genseq/AAL1982.DAT.*
5: /S1D58/gcgdata/genseq/genseq/AAL1983.DAT.*
6: /S1D58/gcgdata/genseq/genseq/AAL1984.DAT.*
7: /S1D58/gcgdata/genseq/genseq/AAL1985.DAT.*
8: /S1D58/gcgdata/genseq/genseq/AAL1986.DAT.*
9: /S1D58/gcgdata/genseq/genseq/AAL1987.DAT.*
10: /S1D58/gcgdata/genseq/genseq/AAL1988.DAT.*
11: /S1D58/gcgdata/genseq/genseq/AAL1990.DAT.*
12: /S1D58/gcgdata/genseq/genseq/AAL1991.DAT.*
13: /S1D58/gcgdata/genseq/genseq/AAL1992.DAT.*
14: /S1D58/gcgdata/genseq/genseq/AAL1994.DAT.*
15: /S1D58/gcgdata/genseq/genseq/AAL1995.DAT.*
16: /S1D58/gcgdata/genseq/genseq/AAL1996.DAT.*
17: /S1D58/gcgdata/genseq/genseq/AAL1997.DAT.*
18: /S1D58/gcgdata/genseq/genseq/AAL1998.DAT.*
19: /S1D58/gcgdata/genseq/genseq/AAL1999.DAT.*
20: /S1D58/gcgdata/genseq/genseq/AAL2000.DAT.*
21: /S1D58/gcgdata/genseq/genseq/AAL2001.DAT.*
22: /S1D58/gcgdata/genseq/genseq/AAL2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6727	98.7	1404	13	AAK26049	MSF precursor. Sy
2	6727	98.7	1404	22	AAK60568	Human megakaryocy
3	6727	98.7	1404	22	AAK29773	Human megakaryocy
4	6151	90.3	1299	22	AAK24322	Human EST encoded
5	3484	51.1	902	22	AAK29778	Human MSF-derived
6	1707.5	25.1	472	22	AAK60569	Bovine MSF ortholo
7	1654	24.3	452	16	AAK80041	Human megakaryocy
8	1171	17.2	5179	22	AAK24516	CG997 predicted an
9	981	13.4	763	21	AAK38942	Arabidopsis thalia
10	950	14.9	1664	19	AAK43106	C. thermocellum O
11	776.5	11.4	1325	22	AAK03645	Peptide #2327 encc

12	715	10.5	763	18	AAW18852
13	627	9.2	4412	18	AAW133666
14	612	9.0	572	18	AAW138555
15	547.5	8.0	804	7	AAW60570
16	542	8.0	847	21	AAW54466
17	520.5	7.6	788	21	AAW54466
18	508	7.5	1937	21	AAW117222
19	506.5	7.4	744	9	AAW82975
20	496.5	7.3	2971	21	AAW41233
21	489.5	7.2	2972	22	AAW50366
22	489.5	7.2	3118	22	AAW50366
23	489	7.2	1721	21	AAW117222
24	488.5	7.2	826	13	AAW2604
25	488	7.2	617	22	AAW0418
26	488	7.2	617	22	AAW0418
27	488	7.2	957	21	AAW59288
28	488	7.2	957	22	AAW24511
29	488	7.2	1721	19	AAW48299
30	485	7.1	1127	22	AAW9554
31	476.5	7.0	652	9	AAW82974
32	467.5	6.9	511	22	AAW1488
33	467.5	6.9	511	22	AAW2751
34	467.5	6.9	511	22	AAW0260
35	455.5	6.7	750	20	AAW0547
36	450.5	6.6	378	12	AAW1416
37	446.5	6.6	378	12	AAW1416
38	443	6.5	2870	21	AAW9555
39	442.5	6.5	3178	21	AAW9555
40	442.5	6.5	751	16	AAW8003
41	437.5	6.4	2819	22	AAW3540
42	429	6.3	1610	21	AAW82250
43	429	6.3	1610	21	AAW7964
44	428.5	6.3	2655	22	AAW1453
45	428.5	6.3	2655	22	AAW26505

## ALIGNMENTS

Mycobacterium tuberculosis  
 Sequence g1101742  
 Mycobacterium tuberculosis  
 Sequence of the Pa  
 Amino acid sequenc  
 Amino acid sequenc  
 Cryptosporidium pa  
 Bioadhesive precu  
 Human ORFX ORF95  
 Human SRAP. Homo  
 Human SRAP. Homo  
 P. yoshii SS2 ant  
 Peptide #2892 enco  
 Peptide #2865 enco  
 Human MUC11 polype  
 G900P predicted am  
 Cryptosporidium pa  
 Human protein sequ  
 Bioadhesive precu  
 Peptide #1317 enco  
 Peptide #1349 enco  
 Peptide #1289 enco  
 C. albicans Koli P  
 PRP 37b. Trillicum  
 PRP encoded by clo  
 Caenorhabditis eleg  
 Caenorhabditis eleg  
 Japanese sea musse  
 Human 07C927 gene  
 EYFP-DYKD-NAP1-EBFP  
 Bifunctional caspase  
 Peptide #967 enco  
 Peptide #967 enco

RESULT	1
AAAR26049	
ID	AAAR26049 standard; Protein; 1404 AA.
XX	
AC	AAAR26049;
XX	
DT	02-FEB-1993 (first entry)
XX	
DE	MSF precursor.
XX	
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF
KW	stability; proteolytic cleavage; adhesion; alternative splicing.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Region
FT	1..26
FT	/label= Exon_I
FT	26..67
FT	/label= Exon_II
FT	67..107
FT	/label= Exon_III
FT	107..157
FT	/label= Exon_IV
FT	157..200
FT	/label= Exon_V
FT	200..1141
FT	/label= Exon_VI
FT	1411..1166
FT	/label= Exon_VII
FT	1166..1212
FT	/label= Exon_VIII
FT	1213..1266
FT	/label= Exon_IX

Mon Apr 29 08:35:31 2002

us-09-556-246-1\_copy\_200\_1140.rag

/b. 2 - 3

Page 16









```

QY 132 PTPKAEKTTGSPALTTPEKPTPTTPKEPASTTPKEPTTITKSAPTTPKEPAPTTPK 191
DB 3 PVP-----APRALPRLPAPPAEPKSKPFPAP-----PAPCCMLVSAAP 46
QY 192 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKE 251
DB 47 PCP---PAP---PAPKPKSKAPFPVPAPAPARELAPLP---PAP-----PAPRE 90.
QY 252 PAPTPKEPTPTT-----PKEPAPTTPKEPAPTTPK-EPAPTPK-KPAPTTPKEPAPTTPK 305
DB 91 SRPLAPCPPPPVVLPDPPEAPAPVPAPAPSPFPFPAPKFPAPVPV---PVPNSPP 148
QY 306 EP---APTTPKEPPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAP 363
DB 149 FPFIPAPALPPAP-----PAPPLANSPLPPAPLPAGT---PPAEPWPVPAAPKSKPA 201
QY 364 TTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAP-----TTTKPAPTAKKEPAPTTPKETAP 419
DB 202 SPPTPPAP-----PMAPLMEFPPLPVPVPDPLSKETCPAPAPPLPAPAPVLPVPPLP 256
QY 420 TTPKRLTPPTPEKLAP-----TTPKEPAPTTPPEELAPPTPEEPTTPPEEPAPTTP-- 469
DB 257 PVPKLIAPPAP---APPVAVAAVLPACPPPLPPLPNHPAPAPAPVPVPVLPAPLP 313
QY 470 ---PKAAPPTPKEPAP----- 483
DB 314 PAPAPAPVPVLPAPLPISGRVAVVWKSFTLISFCRCVSGVLAGALNPSRPSPL 373
QY 484 ---TTPKEPAPTTPKEPAPTTPKETAP-----PTTPKGTAP---PTTLKEAPAPTTP-KKAPKE 533
DB 374 TTTPLAPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLPAPLP 431
QY 534 LAPPTTKEPTSTSDKAPAPTTPKGTAPAPTTPKEPAP-----PTTPKEPAPTTPKGTAPPTL 587
DB 432 ---PAPKPKVPTLPPLAPAPPEK-TVPLPPGSCCPSEKPNPAPPEPEPKSSPALP 488
QY 588 KEAPAPTTP---KKPAPKELAPPTTKGP-TSTSDKAPAPTTPKET---APTTPKEPAPTTP 640
DB 489 APPAPAPASAVVPAPPLPAPPAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 548
QY 641 KKAPAPTTPPTPTTSEVSTPTTKPTTKPTTKHNSP---DBSTPELSAEPKPA--LENSP 694
DB 549 ---PAPAPAPLPAPPLS---APPPCPVPAPAPPTTPAPAPAPAPAPAPAPAPAPAPAPAP 604
QY 695 KEGCVPTTKTPAATKPEMTTAAKDKTTERDLKPTPETTTAAPKMKETATTEKTESKI 754
DB 605 ---PAPAPAPAPAPALPVPNPA-----PPLPAAPK-----SPAL 637
QY 755 TATTTQVSTTQDTTPFKITTLTKITTLAKVTTTKTTTTELMKPEETAKPKDRATN 814
DB 638 PAAPAPAPAPAPVATCP-----PPAPAPAPAP 665
QY 815 SKAATTPKOKP---TKAPKPTSTKKPKMTMPVRRKPTTPPRKMTSTV---PELNP 865
DB 666 SMLAPAPAPAPPLPPLATPPAPAPAPPLPMSAPAPPLPAPAPAPAPPLPAPAPPLPAPAPPLP 725
QY 866 TSRIAEMLQTTTRPMOTPNKSLVEVNPXSEDAAGAGET 905
DB 726 VPAPAPLAPLPINGRPVFAKXNSLI-----GSSSGAT 756

```

RESULT 12

```

AA53666
ID AA53666 standard; Protein; 4412 AA.
AC AA53666;
XX
XX
DT 22-FEB-2000 (first entry)
DE Sequence g1/1017427/emb/CNA62189 from an alignment with protein 608.
KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
KW bone development; g1/1017427/emb/CNA62189.

```

```

XX OS Unidentified.
XX PN W09960164-A1.
XX PD 25-NOV-1999.
XX PF 14-MAY-1999; 99NO-US11066.
XX PR 15-MAY-1998; 98US-0085673.
XX PA (QUAR-) QUARK BIOTECH INC.
PI Elnat P, Mor O, Skalter R, Feinstein E, Faerman A;
XX WPI; 2000-053304/04.
XX PT Identification of stress induced genes for determining risk and
XX PT preventing, treating or controlling osteoporosis
XX PS Claim 32; Fig 6A-R; 308pp; English.
XX CC The present sequence is obtained from a clustrel X alignment with
XX CC protein 608. Protein 608 was identified using the method of the invention
XX CC after subjecting rat osteoblasts to mechanical stress. Expression of the
XX CC 608 gene was found to be upregulated by about 3-fold in cells subjected
XX CC to mechanical strain. The specification describes a method for the
XX CC identification of genes responsive to a specific mechanical stress. The
XX CC method comprises applying the mechanical stress to an organism (tissue
XX CC or cells comprising bone cells), isolating the specific cellular
XX CC mRNA in comparison with control samples. The method is used to identify
XX CC genes whose expression is responsive to a specific stress. The identified
XX CC genes are employed in determining risk associated with a physiological or
XX CC disease state. The risk determining methods are used for testing a
XX CC medicament for gene therapy. These medicaments, or genes identified by
XX CC the method of the invention, are used for treating, preventing or
XX CC controlling a physiological or disease state (especially osteoporosis or
XX CC bone density or other factors causing or contributing to osteoporosis or
XX CC its symptoms or other conditions involved in mechanical stress or its
XX CC lack. The methods can also be used for advancing research or studies in
XX CC bone development.
XX SQ Sequence 4412 AA;

```

```

Query Match 12.5%; Score 625.5; DB 21; Length 4412;
Best Local Similarity 26.2%; Pred. No. 5.2e-27;
Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;

```

```

QY 99 TDGKEKTTSAKETSIEKT-----SAKDLAPTSVLAKPTPKAETTKGPAALT----- 146
DB 3515 tdtirgllqrvceehrvkhrvlefeaevefekpkapp-----kypisaskllp 3566
QY 147 TPKETPTTP-KEPASTTPKEPTTITKSAPTTPKEP-APTTPKSAPTTPKEPAPTTPKE 204
DB 3567 pkkpklkvprkpkpavkvpkklivveekvrvpckvrvpckvrvpckvrvpckvrvpckvrvpck 3621
QY 205 PAPTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPK---PAPTPKEPAPTTPKEP 260
DB 3622 ---ekkvvpapkpap-----pkkvpeapkvpekvvpkpkkvpckkvpevp 3672
QY 261 TPTTPKEPAPTTPKEPAPTTP-----KEPAPAPKPPAPPTTPKEPAPTTPKEPAPTTPK 313
DB 3673 kaavpekavpealpkpsepppevfepespsapkp-kpavpvrvpvpkavpekvv 3731
QY 314 EDSPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKESPTTPKEPAPTTPKEPAPTTP 373
DB 3732 paap--pkkpvtvpk-vpeapkvpekkvvp-pkkpvpvpck-----vpevpkvav 3782
QY 374 PKKPAAPTTPKEPAPTTPKAPAPTTPKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 429
DB 3783 pekvp-----ealpkpsepppevfepespsapkp-kpavpvrvpvpkavpekvv 3838

```







Db 2360 pltttttvcptcgtc---tqtp--ttpttttttvc 2392

RESULT 8

AAG38942

ID AAG38942 standard; Protein: 763 AA.

XX

AC AAG38942;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PE 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135258.

PR 24-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 16-JUN-1999; 99US-0139454.

PR 17-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

---

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139839.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140635.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145089.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145293.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148319.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

AAM24516  
 ID AAM24516 standard; Protein; 5179 AA.  
 XX  
 XX AAM24516;  
 XX DT 12-OCT-2001 (first entry)  
 XX DE C899P Predicted amino acid sequence.  
 XX  
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 KM immunogenic; gene therapy; vaccine; colonic cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200149716-A2.  
 PD 12-JUL-2001.  
 XX  
 PF 29-DEC-2000; 2000WO-US35596.  
 XX  
 PR 30-DEC-1999; 99US-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jiang Y;  
 XX  
 WP1: 2001-441847/47.  
 XX  
 PT Colon tumor associated proteins and nucleic acids useful for the  
 PT prevention, diagnosis and treatment of colonic cancer -  
 XX  
 Claim 2; Page 446-462; 472pp; English.  
 XX  
 PS The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patients own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA1284460 to AA129512  
 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences  
 XX given in the exemplification of the present invention.  
 XX  
 Sequence 5179 AA;  
 SQ

Query Match	23.3%	Score 1168;	DB 22;	length 5179;
Best Local Similarity	32.5%;	Pred. No. 2.3e-57;		
Matches 363;	Conservative 53;	Mismatches 448;	Indels 254;	Gaps 40;
21 DEAGSGLDNGDFK-----	VTPPTSTTQH-NKVSNSPK-----			52

Dh	1312	dhpsgsddgdteprfgvgagapedlectrsvxkdlhslteghqkwqcdvsvglcknedqf	1371
Qy	53	-----ITAKPINPNSLPPNSDTSKEKSLVNNKTEVYRK	88
Dh	1372	gn9pfglcydykIrvncwpmdkciltspstpsppsttcllplrttspstt----	1427
Qy	89	ETTTHKKOISJUDGKEKTEKTSAKETQSIIEKUSAKDLAPTSKVLAKEPTAEETTHKGPALTT	148
Dh	1428	-tttpprttspsttspsttlttpp-----lptc-----tspsttsttpprtt	1470
Qy	149	KEPT-----PTPKPEASTTPKEPPTTIKSAP--TTKEP-----APTTISAP--TTP	194
Dh	1471	spsttspsttspsttspsttcltpprttspsttspmmtcllplpasccllplrttspstt	1530
Qy	195	KEAPATTPKKEAPATTPKEB-----APTPKKEAPATTPKSAP--TTPKKEAPATTPKKA	245
Dh	1531	ttpprttspsttpprttpprtstcllplrttspsttcltpprttspsttspsttspstt	1590
Qy	246	PTPKPEAPATTPKEPPTTPKKEAPATTPKKEAPATTPKEB-----APTAPEAPATTPKE	298
Dh	1591	tttpprttspsttspsttcltpprttspsttpprttpprtstcllplrttspsttclt	1650
Qy	299	PAPTPKKEAPATTPKPSPTTPKKEAPATTPKSAPTTPKKEAPATTPKSAAP--TTPKPS	355
Dh	1651	pprttspsttspsttspsttcltpprttspsttspsttspsttcltpprttspsttspstt	1709
Qy	356	TTTTKEAPATTPKPEAPATTPKPKAPATTPKKEAPATTPKKEAPATTPKKAAPTAKKEAPATTP	414
Dh	1710	tttspsttspsttspsttmttspsttspsttcltpprttspsttcltpprttspsttcltpprtt	1769
Qy	415	-----KETAPTTP-----KULTP	427
Dh	1770	tsptsttpprttpprttclpvcnlcnwlgwldsgkpnfhkpggdtecllgdvcpbgpwaaniscratmtp	1829
Qy	428	-----TTPKEAPATTP	437
Dh	1830	dvpiqglqgtwcvdsvglcknedqbgvgvlpmafcInyngvcecvrtqblmttclt	1889
Qy	438	PEKPAPTTPBELAPTPPEEBPTTP--TPEEAPATTPKKAAPNTPKKEAPATTPKEP---AP	491
Dh	1890	temppttpprttcltcltcltpprttpprttpprttpprttcltcltcltpprttpprttpprtt	1949
Qy	492	TTPEKEAPATTP--TPEKETAATTPPKGAPATTPKKEAPATTPKPKAPAKELAPTTKPESTGDKP	550
Dh	1950	tttpprttpprttpprttpprttcltcltcltpprttpprttpprttpprttcltcltpprttpprtt	2009
Qy	551	AP--TTPKGAPATTPKKEAPATTPKKEAPATTPKKGATTPKKEAPATTPKPKAPAKELAPTTK	609
Dh	2010	ttpttpprttcltcltpprttpprttpprttpprttcltcltcltpprttpprttpprttcltclt	2068
Qy	610	GPSTGSDKAP--TTKEKETAATTPKKEAPATTPKKAAPATTPPEPPTTPSTSEVSTPTTKPEPT	668
Dh	2069	ttpprttpprttpprttpprttcltcltcltpprttpprttpprttcltcltcltpprttpprttpprtt	2127
Qy	669	TIHKSPESPELSEAEPTPALENSKEP-----GVPT--TKTPAAKPEMTTJAK	717
Dh	2128	t-----tpprttcltpprttpprttpprttpprttcltcltcltpprttpprttpprttcltclt	2184
Qy	718	DKTERDLR--TTP--ETTTPAAKMT--KETATTEKTESKTIATTOVYSTTQDT	769
Dh	2185	pprttpprttpprttpprttcltcltcltpprttpprttpprttpprttcltcltcltpprttpprtt	2243
Qy	770	TPKTIITLTKTTTLAPKVTYT--KKTITTTBELMMKPEETAKPKORATNSKATPKPKQKPKKA	828
Dh	2244	tp--ttt--tttpprttpprttpprttpprttcltcltcltpprttpprttpprttpprttcltclt	2299
Qy	829	P-KKPTSTKKPKTUPKRVKKTTPTPRKMTSTMPELNP---TSRIAEAMLQTTTP--PNQ	883
Dh	2300	pprttpprttpprttpprttcltcltcltpprttpprttpprttpprttcltcltcltpprttpprtt	2359
Qy	884	PNSKLVEPNKPSDAGABSETPHMLLRPHVFMPEVTP	921

D6 ptprrptgqrpttprttttvrcprrpqrgqrrtttttttcvrrprrpgtqrptt 2309  
QY 884 PNRKLVEYNPKSEDAGACBGETPHMLLRPHVMEPTVP 921

```

Db      61 prpslprnsdsketelvtvketvctttnkgtstcdgkelttsaktsiektask 120
Oy      121 DLAPTSKVLAKPPEAKETTTKGPAALTTPEKEPTTPPKPASTTPKEPTTTTKSAPTPK 180
Db      121 dlapskvlakpckpaeltckgpa1-----tckpcasttkpckptcktsakptpk 172
Oy      181 EPAPTTTKSAPTPPEAPTTTKPEAPTTTPPKPAPTTTKEAPTTTKSAPTPKPAPTT 240
Db      173 epapttkcaptkpkeap-ctckepapt-kepap-ctckepaptckepapt-kepapct 228
Oy      241 PKKPAPTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAP 300
Db      229 -kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-ke 281
Oy      301 PTPPEAPTTTPKEAPTTTPKEAPTTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360
Db      283 ptt-kepap-ctckepapt-kepapctckepap-ctckepaptckepapt-kepap-ct 335
Oy      361 PAPPTPKPEAPTTTPPKPAPTTPKEAPTTTPKEAPTTTKKAPAPAPKPEAPTTPEK 420
Db      336 paptt-kepap-kepap-kepap-kepap-kepap-ctckepapt-kepap-kepap 388
Oy      421 TPXKLTPTPEKLAFTPEKAPPTPEELAPTTPEEPPTTPPEEPAPTTPKAAADNTPKE 480
Db      389 t-kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-ke 440
Oy      481 PAPPTPKPEAPTTTPKEAPTTTPKEAPTTTPKGATPTTKPEAPTTTPKPKKAPKELAP 540
Db      441 paptt-kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-ctk 488
Oy      541 EPTSTSKAPAPTPPKGAPTTPEAPTTTPKEAPTTTPKGAATPTTKPEAPTTTPKPKAP 600
Db      489 epaptke-kepap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-ke 541
Oy      601 KELAPTTKGATSTSDAPTPPEKAPTTTPKEAPTTTPKPKKAPPTPEPTTSEVST 660
Db      542 -----tckepaptke-kepap-kepap-kepap-kepap-kepap-kepap-kepa 589
Oy      661 PTTKEPTTIKSPDESTPELSAEPTRKALNSPEKPGVPTTKTPAATKPEMTTAKDKT 720
Db      590 -ptckepaptckepapt-kepap-kepap-kepap-kepap-kepap-kepap-kepa 636
Oy      721 TERDRTPTETTAAPKKTETATTTETKTSKITATTTQVTSSTTQDTTPFKTTLKT 780
Db      637 -----pap-kepap-kepap-kepap-kepap-kepap-kepap-kepap-ke 667
Oy      781 TLAPKVTITTKTITTEIMNKPETAKPKDRATNSKATTPPKAPKPKPTSTKKPKT 840
Db      668 -----tckepaptckep-apt-ctckep- 686
Oy      841 MPVRKKEPTTPPKKMTSTMBELNPTSRIAEAMLQTTTRPNQNSKLVENPKSEDAGG 900
Db      687 -----apctcpckmstmpelnpstarielaaml-tttrpqrpsnklvempkseda 739
Oy      901 AEGEPHMLLRHYEMPEVTPDMQYDLPRVNOGIINPMLS 941
Db      740 aegcpmlilprhvfmpcvcpdmqydlprvnpqililnplms 780

```

```

XX      XX Homo sapiens.
OS
XX      Key Location/Qualifiers
FH      MISC-difference 393..396
FT      MISC-difference 444..446
FT      MISC-difference 444..446
FT      MISC-difference 444..446
XX      W09523861-A1.
XX      08-SEP-1995.
XX      06-MAR-1995; 95MO-CN00015.
XX      04-MAR-1994; 94CN-0112066.
XX      (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.
XX      Gu X, Han Z, Shen Q;
XX      WPI; 1995-320576/41.
XX      DR N-PSDB; AAT04546.
XX      New haematopoietic cell growth factor - used for treating
XX      thrombocytopenia and hematocytopenia
XX      Example; Page 23; 36pp; Chinese.
XX      This sequence represents the human megakaryocytopoietin (MPO) protein.
XX      This sequence was purified using a carrier which can couple wheat germ
XX      agglutinin and heparin to separate MPO. Fragments of this sequence (see
XX      AAR80039 and AAR80040) were used to produce the amplification primers
XX      shown in AAT04344 and AAT04345. The fragments amplified by these primers
XX      can then be used as probes to screen human cDNA libraries for MPO cDNA.
XX      The MPO cDNA can then be inserted into a plasmid which is used to
XX      transform cells to produce MPO. The MPO sequence is capable of promoting
XX      colony formation of megakaryocytes, enlarging the size of megakaryocytes
XX      and stimulating the proliferation of multipotential stem cells. The
XX      factor may be used for treating thrombocytopenia and hematocytopenia.
XX      The purification method can be used to isolate MPO from human urine or
XX      sera of patients with aplastic anaemia, and from animal blood or urine
XX      by radiation exposing the animals to induce aplastic anaemia.
XX      Sequence 452 AA:
SQ
Query Match 27.1%; Score 1357.5; DB 16; Length 452;
Best Local Similarity 88.3%; Pred. No. 4.8e-69;
Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;
Oy      1 VKDKKRRTRKKKPPRPVVDKAGSGLDNDGKVTTPDTSTTQHNKVSPTKTTAKPIN 60
Db      157 vkdkkrrtrkkkpprvvdkagsgldngdkfvttcpdtsltqhnkvsptkltakpin 216
Oy      61 PRSLPNSDTSKETSILTVKKEETTVETKETTNNKOSTGKKEKTSAKETOSTEKTSK 120
Db      217 prslpnsdtsketsiltvketvctttnkgtstcdgkelttsaktsiektask 276
Oy      121 DLAPTSKVLAKPPEAKETTTKGPAALTTPEKEPTTPPKPASTTPKEPTTTTKSAPTPK 180
Db      277 dlapskvlakpckpaeltckgpaaltckpckptcktsakptcktsakptpk 336
Oy      181 EPAPTTTKSAPTPPEAPTTTKPEAPTTTPPKPAPTTTKEAPTTTKSAPTPKPAPTT 240
Db      337 epapttkcaptkpkeap-ctckepapt-kepap-ctckepaptckepapt-kepapct 396
Oy      241 PKKPAPTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAP 298
Db      397 ctcp-----tckpckpplprslhprptckepaptckepapapckpckpplple 443

```

RESULT 7

CC human MSF.  
XX  
SQ Sequence 1404 AA;

Query Match 100.0%; Score 5011; DB 22; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 1.6e-273;  
Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VKDNKKNRRTKKKKPPKPPVVDGAGSGIDNGDFKVTPTPTSTTOHKKVSTSPKITTAKPIN 60
DB 200 vkdnkknrtkkkppkppvvdagsgjdngdfkvtptdstlqhnkfstapklitakpin 259
QY 61 PRPSLPNSDTSKETSILVVKETVETKETTNTKOSTDGKEKTSAKETOSIEKTSAK 120
DB 260 prpslpnsdtsketsilvvnkettvetkettlnkstlqgkelttsaketqsiektak 319
QY 121 DLAPTSKVLAKPPKPAETTTKSPALTPPKETPTTPKEPASTTPKEPTPTTIKSAPTPPK 180
DB 320 dlaptskvlnkppkpaetttkspaltppkettpttpkepasttpkeptpttiksaptppk 379
QY 181 EPAPTTTKSAPTPPKKEPAPTTTKKEPAPTTTKKEPAPTTTKSAPTTTKSAPTTKEPAPT 240
DB 380 epaptttksaptppkkepapttkkepapttkkepapttkkepapttksaptppkaptp 439
QY 241 PKKAPPTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKAPAPTTKEPA 300
DB 440 pkkapptpkkepapttkkepapttkkepapttkkepapttkkapaaptpkapa 499
QY 301 PTPKKEPAPTTKEPSPPTPKKEPAPTTTKSAPTTTKSAPTTTKSAPTTKEPSPPTTKE 360
DB 500 ptpkkepapttkespptpkkepapttkksaptttkkepapttksaptpksppttkke 559
QY 361 PAPPTPKKEPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPT 420
DB 560 papptpkkepapttkkapaaptptkkapaaptptkkapaaptptkkapaaptpk 619
QY 421 TPBKLTPTPEKLAPTPEKAPAPTTPEELAPTTPEEPPTTPPEEPAPTTPEEAPNTPKPE 480
DB 620 tpbkltptpeklaptpekapaaptppeelaptppeeppttppeepaptpkpeapntpk 679
QY 481 PAPPTPKKEPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPTTTKKAPAPT 540
DB 680 papptpkkepapttkkapaaptptkkapaaptptkkapaaptptkkapaaptpk 739
QY 541 EPTSTTSKAPAPTTPKGAPAPTTPKGAPAPTTPKGAPAPTTPKGAPAPTTPKGAPAPT 600
DB 740 eptsttskapaptpkgaaptpkgaaptpkgaaptpkgaaptpkgaaptpkgaap 799
QY 601 KELAPTTTKGPTSTSDKAPAPTTKEPAPTTPKKAPAPTTPEPPTPTSEVST 660
DB 800 kelpaptttkgptstsdkapapttkepapttkkapaptpkpaaptpkpaaptpk 859
QY 661 PTTTKEPTTIHKSPEDESPELSAETPKALNSKREPEVPTTKPAAKKPEMTTAKAKXT 720
DB 860 ptttkepttihkspepespelsaetpkalnskrepevpttkpaakkpemttaakxt 919
QY 721 TERDRTPEPTTTAPKMTKETATTEKTESKITATTQVSTTQDTPTFFKITTLKTT 780
DB 920 terdrtpeptttapkmtketattektekskitattqvsttqdtptffkitllkt 979
QY 781 TLAPKVTYTTKKTITTTETIMNKRPEETAKPKDRATNSKATTPKQKPTKAPKKPSTSKAKXT 840
DB 980 tlapkvtyttkktitttetimnkrpeetakpkdratnskattpkqkpkakpkpstkakp 1039
QY 841 MPRVAKKPTTTPPKMTSTMPRLNPTSRILAEMLOTTTRPNQTPMSKIVLVKPKSEDDGG 900
DB 1040 mprvakp ttppkmtstmp rlnptsrilaemlo ttrpnqtpmskivlvkpkse d d g g 1099
QY 901 AEGEPHMLRPHVMPPEVTPDMOYLPRVNOGIITNPMLS 941
DB 1100 aegep hmlrphvmppe vtpdmoylprvno giitnpmls 1140

```

RESULT 5  
AAB29778  
ID AAB29778 standard; protein; 902 AA.  
XX

AC AAB29778;  
XX  
DF 28-FEB-2001 (first entry)  
XX

DE Human MSF-derived tribonectin.

XX Human tribonectin; MSF; megakaryocyte stimulating factor;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KW friction coefficient reduction; gene therapy; antiarthritic;  
KW osteopathic.

XX Homo sapiens.  
OS  
PN W0200064930-A2.  
XX

PD 02-NOV-2000.

PF 24-APR-2000; 2000MO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PI Jay GD;

XX WPI; 2001-024673/03.

PT Novel tribonectin polypeptide useful as lubricant for treating  
XX osteoarthritis, comprises O-linked lubricating moiety -

PS Disclosure; Fig 1; 47pp; English.

XX The invention relates to a human tribonectin which is a product of  
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonectin has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76  
CC repeats of a motif having at least 50% identity to the sequence KEAPPT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a  
CC human MSF-derived tribonectin; a biocompatible composition comprising a  
CC human tribonectin for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC tissues of a mammal, where the injury is caused by a surgical insertion  
CC of trauma, or an artificial device e.g., an orthopaedic implant. In  
CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonectin may be used in gene therapy. The present sequence represents  
CC a substantial portion of a human MSF-derived tribonectin.

XX Sequence 902 AA;

Query Match 56.1%; Score 2812.5; DB 22; Length 902;  
Best Local Similarity 71.4%; Pred. No. 2.3e-150;  
Matches 672; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

```

QY 1 VKDNKKNRRTKKKKPPKPPVVDGAGSGIDNGDFKVTPTPTSTTOHKKVSTSPKITTAKPIN 60
DB 1 vkdnkknrtkkkppkppvvdagsgjdngdfkvtptdstlqhnkfstapklitakpin 60
QY 61 PRPSLPNSDTSKETSILVVKETVETKETTNTKOSTDGKEKTSAKETOSIEKTSAK 120

```





PT immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1: 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine preceding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC sequences derived from human megakaryocyte colony stimulating factor  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC primary transcript of this gene may be cleaved in different ways to  
 CC yield a family of mRNA's each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor i.e. one or more of these exons may contain  
 CC sequences which direct proteolytic cleavage, adhesion, organisation of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.

XX Sequence 1404 AA:

SQ

Query Match 100.0%; Score 5011; DB 13; Length 1404;

Best Local Similarity 100.0%; Pred. No. 1.6e-273;

Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTOHNNKYSTSPKITTAKPIN 60  
 DB 200 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTOHNNKYSTSPKITTAKPIN 259  
 QY 61 PRSLPNSNSTSKRTSLTVNKETVETKETTNNKORTSGKKEKTSAAKESOSTEKSATK 120  
 DB 260 PRSLPNSNSTSKRTSLTVNKETVETKETTNNKORTSGKKEKTSAAKESOSTEKSATK 319  
 QY 121 DLAPTSKVLAKPPKAEETTKGSPALTTTPKEPTTPKEPASTPKEPTTTIKSAPTPK 180  
 DB 320 DLAPTSKVLAKPPKAEETTKGSPALTTTPKEPTTPKEPASTPKEPTTTIKSAPTPK 379  
 QY 181 EPAPPTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTSAPTTTPKEPAPT 240  
 DB 380 EPAPPTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTSAPTTTPKEPAPT 439  
 QY 241 PKKRAPTPKRAPATTPKEPTTPKEPAPTTTKEPAPTTTKEPAPTTAPKRAPATTPKEPA 300  
 DB 440 PKKRAPTPKRAPATTPKEPTTPKEPAPTTTKEPAPTTTKEPAPTTAPKRAPATTPKEPA 499  
 QY 301 PTPKRAPATTPKESPTTPKEPAPTTTKSAPTTTKEPAPTTTSAPTTTPKEPAPT 360  
 DB 500 PTPKRAPATTPKESPTTPKEPAPTTTKSAPTTTKEPAPTTTSAPTTTPKEPAPT 559  
 QY 361 PAPPTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPAT 420  
 DB 560 PAPPTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPAT 619  
 QY 421 TPKKLTPTPEKLAATTPPEKAPATTPPEELAPATTPPEEPTTPPEEAPATTPKAAAPNPEKE 480  
 DB 620 TPKKLTPTPEKLAATTPPEKAPATTPPEELAPATTPPEEPTTPPEEAPATTPKAAAPNPEKE 679  
 QY 481 PAPPTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPAT 540  
 DB 680 PAPPTTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPATTPKRAPAT 739

QY 541 EPTSTSDKAPATTPKGNATTPKKEPAPTTTPKEPAPTTTPKGNATTPKKEPAPTTTPKKNAP 600  
 DB 740 EPTSTSDKAPATTPKGNATTPKKEPAPTTTPKEPAPTTTPKGNATTPKKEPAPTTTPKKNAP 799  
 QY 601 KELAPATTPKGNATTPKKEPAPTTTPKEPAPTTTPKKEPAPTTTPKKNAPPTTPPTTSEVST 660  
 DB 800 KELAPATTPKGNATTPKKEPAPTTTPKEPAPTTTPKKEPAPTTTPKKNAPPTTPPTTSEVST 859  
 QY 661 PPTTPKEPTTIKHSPPDESTPELSAEPTPKALBNSPKEPGYPTTKTPAANKPEWTTAKOKT 720  
 DB 860 PPTTPKEPTTIKHSPPDESTPELSAEPTPKALBNSPKEPGYPTTKTPAANKPEWTTAKOKT 919  
 QY 721 TTRDRTTPPETTTAPKMTKEATTTTEKTESKITTATTOVSTTODTPPKITTLKTT 780  
 DB 920 TTRDRTTPPETTTAPKMTKEATTTTEKTESKITTATTOVSTTODTPPKITTLKTT 979  
 QY 781 TLAPKVTYTKTITTTTEINMKPEETAKPKDRATNSKATTPKQKPTKAPKKTSTKKEPT 840  
 DB 980 TLAPKVTYTKTITTTTEINMKPEETAKPKDRATNSKATTPKQKPTKAPKKTSTKKEPT 1039  
 QY 841 MPVVRKPKTTTPPKMTSTMPPEINPTSRJAEMLOTTTPRNOPTNSKILVEVNPKSDAGC 900  
 DB 1040 MPVVRKPKTTTPPKMTSTMPPEINPTSRJAEMLOTTTPRNOPTNSKILVEVNPKSDAGC 1099  
 QY 901 AEGETPHMLRPHEVPEVTPOMDYLPARYNOCIIINPMLS 941  
 DB 1100 AEGETPHMLRPHEVPEVTPOMDYLPARYNOCIIINPMLS 1140

RESULT 3

AAB60568 AAB60568 standard; Protein; 1404 AA.

AC AAB60568;

DT 27-APR-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF, CACP).

XX Human: CACP protein; campodactyly-arthropathy-coxa vara-pericarditis;

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW Chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

OS Homo sapiens.

XX WO200107068-A1.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000MO-US20002.

PK 23-JUL-1999; 99US-0145328.

PR 19-JUL-2000; 2000US-0145328.

PA (UYCA-) UNIV CASE WESTERN RESERVE.

PI Warman ML;

DR WPL; 2001-182721/18.

PT New composition comprising the campodactyly-arthropathy-coxa  
 PT vara-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints

PS Example 1; Page -: 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactyly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous

DR N-PSDB: AAH98981.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 XX  
 PS Claim 20: Page 1198-1201: 1275pp, English.  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 XX Sequence 1299 AA;

Query Match 100.0%; Score 5011; DB 22; Length 1299;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-273;  
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQDNKKNRKPKPPVNDAGSLNGDFKVTPTSTTQHNKYSTSEKITTAKPIN 60  
 DB 200 vdknknkntckkppkppvndaagslngdfkvtptsttqhnkystspkittakpin 259  
 QY 61 PRPSLPNSDTSKETSILTANKETVETKETTNNKOTSDGKEKTSKESQSTESAK 120  
 DB 260 prpslpnadskselsilvnhkettvettltnkqstsdgketsketskqstksak 319  
 QY 121 DLAPTSKVLAPTKAETTTGPAITTPKEPTPTPKKPASTTPEKPTTIKSAPTTPK 180  
 DB 320 dlapskvlaptkpkaetltgpaaittpkepttpkpkpastaatpkepttlksaptpk 379  
 QY 181 EPAPTTKASAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAPTT 240  
 DB 380 epapttksaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 439  
 QY 241 PKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAP 300  
 DB 440 pkkaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeap 499  
 QY 301 PTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 360  
 DB 500 pttkeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 559  
 QY 361 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 420  
 DB 560 papttpkeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptp 619  
 QY 421 TPKLTPTTPEKLAPTTPKEKAPPTTPPEELAPTTPEEPTPTTPEEAPTTKAAAPNTPK 480  
 DB 620 tpkltpttpeklapttpkekapttpppeelapttpeeptpttpeeapttkkaapntpk 679  
 QY 481 PAPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 540  
 DB 680 papttpkeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptp 739  
 QY 541 EPTSTTSKAPPTPKGAPTTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAP 600  
 DB 740 eptsttsdkapptpkgaptppkkapptpkkapptpkkapptpkkapptpkkap 699  
 QY 601 KELAPTTKGPSTTSKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPPTPKKAPT 660  
 DB 800 kelapttkgspttsdkapptpkkapptpkkapptpkkapptpkkapptpkkap 859  
 QY 661 PTTTKEPTTIHKSPTDSBELSAEPTPKALENPKKPGVPTTKPATPEMTTAKDXT 720  
 DB 860 ptttkepttihksptdsbelaeptpkalennpkpgvpttkpatpemttaakdxt 919  
 QY 721 TERDLTTTETTTAAKMKKETAATTEKTESKITAATTTQVSTTTQDTPPKITTLKTT 780  
 DB 919 terdltttetttaaakmkketaattekteskitaatttqvstttqdtppkittlkt 780

DB 920 terdlrttpeettaapkmktetatttktesskiltattgtvtsltqdltpfkiltlkt 979  
 QY 791 TLAPKVVTTTKKTTTTEITMNRPEETAKPKDRATNSKATTPKOKPTKAPKPKSTKKPKT 840  
 DB 980 tlapkvtttkktttiteitlmnrpeetakpkdratnskattpkpkpstkakpkstkpkpk 1039  
 QY 841 MPVRKPKTTTTPPKKMTSTMBELNPTSKRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAG 900  
 DB 1040 mpvrkpktttppkkmtstmbelnptskriaeamlqtttrpnqtpnsklvevnpksedag 1099  
 QY 901 AEGETPHMLLRPHVMEVPTDMDYLPRVPMQGIITNPMIS 941  
 DB 1100 aegetphmllrphvmevptdmdylprvpmqgiitnpmis 1140

RESULT 2  
 AAR26049  
 ID AAR26049 standard; Protein: 1404 AA.  
 XX  
 AC AAR26049;  
 XX  
 DT 02-FEB-1993 (first entry)  
 XX  
 DE MSF precursor.  
 XX  
 KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
 KW stability; proteolytic cleavage; adhesion; alternative splicing.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..26 /label= Exon\_I  
 FT Region 26..67 /label= Exon\_II  
 FT Region 67..107 /label= Exon\_III  
 FT Region 107..157 /label= Exon\_IV  
 FT Region 157..200 /label= Exon\_V  
 FT Region 200..1141 /label= Exon\_VI  
 FT Region 1141..1166 /label= Exon\_VII  
 FT Region 1166..1212 /label= Exon\_VIII  
 FT Region 1213..1266 /label= Exon\_IX  
 FT Region 1266..1331 /label= Exon\_X  
 FT Region 1331..1373 /label= Exon\_XI  
 FT Region 1373..1404 /label= Exon\_XII  
 FT Region  
 XX  
 XX WO9213075-A.  
 XX  
 XX 06-AUG-1992.  
 XX  
 XX 17-JAN-1992; 92WO-US00433.  
 XX  
 XX 18-JAN-1991; 91US-0643502.  
 XX 10-SEP-1991; 91US-0757022.  
 XX  
 XX (GENY ) GENETICS INST INC.  
 XX  
 XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
 XX WPI, 1992-284660/34.  
 XX N-PSDB; AAQ27223.  
 XX  
 XX New human mega-karyocyte stimulating factors - for treating

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:29 ; Search time 107.17 Seconds  
(without alignments)  
650.396 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1140  
5011  
Perfect score: 1 VKDKKKRRTKKKPPKPPV.....DMDYLPVPNGIINPMLS 941  
Sequence:

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDSB/gcgdata/geneseq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/AA2000.DAT.\*  
21: /SIDSB/gcgdata/geneseq/AA2001.DAT.\*  
22: /SIDSB/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5011	100.0	1299	AAW24322	Human EST encoded
2	5011	100.0	1404	AAW26049	MSF precursor. Sy
3	5011	100.0	1404	AAW60568	Human megakaryocyt
4	5011	100.0	1404	AAW29773	Human megakaryocyt
5	2812.5	56.1	902	AAW29778	Human MSF-derived
6	1357.5	27.1	452	AAW80041	Human megakaryocyt
7	1168	23.3	5179	AAW24516	C899p predicted am
8	981	19.6	763	AAW38942	Arabidopsis thalia
9	942	18.8	1664	AAW43106	C. thermocellum OI
10	771.5	15.4	1325	AAW03645	Peptide #2327 enco
11	715	14.3	763	AAW31852	Mycobacterium tube

12	625.5	12.5	4412	21	AAW53666
13	612	12.2	572	18	AAW31855
14	544	10.9	844	7	AAW60570
15	542	10.8	807	21	AAW54467
16	520.5	10.4	788	21	AAW54466
17	506.5	10.1	744	9	AAW82975
18	496.5	9.9	2971	21	AAW41231
19	493	9.8	1837	21	AAW11726
20	489.5	9.8	2972	22	AAW50363
21	489.5	9.8	3118	22	AAW50362
22	488.5	9.7	826	13	AAW26042
23	488	9.7	617	22	AAW16458
24	488	9.7	617	22	AAW04187
25	488	9.7	957	21	AAW52888
26	488	9.7	957	21	AAW52888
27	485	9.7	1127	22	AAW95411
28	476.5	9.5	652	9	AAW82974
29	467.5	9.3	511	22	AAW14883
30	467.5	9.3	511	22	AAW27312
31	467.5	9.3	511	22	AAW02607
32	466	9.3	1721	19	AAW48299
33	463.5	9.2	1721	21	AAW11727
34	450.5	9.0	378	12	AAW14160
35	446.5	8.9	378	12	AAW14162
36	446.5	8.9	750	20	AAW05477
37	445.5	8.9	472	22	AAW60569
38	443	8.8	2870	21	AAW95559
39	443	8.8	3178	21	AAW95556
40	442.5	8.8	3178	16	AAW80839
41	425.5	8.5	910	22	AAW83007
42	424.5	8.5	1012	20	AAW17406
43	424.5	8.5	1125	21	AAW22934
44	424.5	8.5	1125	21	AAW79637
45	424.5	8.5	1610	21	AAW22870

#### ALIGNMENTS

RESULT 1	AAW24322 standard; Protein; 1299 AA.
ID	AAW24322.
AC	AAW24322.
DF	12-OCT-2001 (first entry)
XX	Human EST encoded protein SEQ ID NO: 1847.
XX	Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	Homo sapiens.
OS	WO200154477-A2.
PN	02-AUG-2001.
PD	25-JAN-2001; 2001WO-US02687.
XX	25-JAN-2001; 2000US-0491404.
XX	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
PA	(HYSE-) HYSEQ INC.
XX	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Gao Y, Drmanac RA, Zhang J, Werhman T;
XX	WPL; 2001-476164/51.

Sequence q1/101742  
Mycobacterium tube  
Sequence of the Fa  
Aminic acid sequenc  
Amino acid sequenc  
Bioadhesive precu  
Human OREF ORF95  
Cryptosporidium pa  
Human SRCAP. Homo  
Human SRCAP. Homo  
P. yoeili SSP2 ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 polype  
C900P predicted am  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
Cryptosporidium pa  
Portion of Cryptos  
ppp 378. Trilicium  
ppp encoded by Cio  
C. albicans Rbt1 p  
Bovine MSF ortholo  
Caenorhabditis ele  
Japanese sea musse  
S. epidermidis ope  
Human atrophin-1 r  
Mouse microtubule-  
Microtubule associ  
EYFP-DEVD-MAP4-EBF













XX WPI: 2001-476286/51.  
 DR Novel single exon nucleic acid probe used to measuring gene expression  
 XX in a human breast -  
 PT  
 XX  
 PS Claim 27; SEQ ID No 12385; 322pp; English.  
 CC  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX Sequence 1325 AA:

Query Match 15.0%; Score 771.5; DB 22; Length 1325;  
 Best Local Similarity 26.0%; Pred. No. 9.1e-36;  
 Matches 302; Conservative 97; Mismatches 466; Indels 295; Gaps 45;

QY 7 NRK-KKPPKP---PVVDAGSGLDNGDFKVTTPDTSTTQ---HNKVSSTPKTT 55  
 Db 60 ntrrhspckpckgnsktlh--stidhheaptsseensngkdpmlnqgrsvdpdst 118  
 QY 56 A-----KPINRPSLPN---SDTSKET-----SITVNKETTVENKET 90  
 Db 119 tlhkasagkhilpapkskincrstgkstvtkrksdktgrpleksmldktsstshkt 178  
 QY 91 TTT-----NKOTSDGKEKTTSAKET-----QSIETK-SADDLAPTSKVLAKPT 133  
 Db 179 ttfsmgsmqtkyqsktsfpekitaasktyktgtprsekedstetvtsadklllktl 238  
 QY 134 PKAETTYKGPALT-----TPKEPPTP---TPKEPASTTPKEPPTPT 170  
 Db 239 knlgetisanelqslaephehgrtanenltspaeptrenatanentltspaepht 298  
 QY 171 TITSA--PTTPKEBAPT-----TTKSAPTTPKKPAPTTTKE---PAPTTKEBAPT 216  
 Db 299 rextanentlapfpaqptenremtanentlfpaeptehgerantentltspaepht 358  
 QY 217 TTKEBAPTTKSAPTTPKEBAPTTPPKKPAPTTTPKEBAPTTPKEBAPT 271  
 Db 359 tane-----ntlpspaephtgerptfandktlssaeptehgerptplanentltspaepht 414  
 QY 272 -TKEPAPTTKEBAPTAPKAPAPTTTPKEBAPTTPKEBAPT---PTTKEBAPTTPKEBAPT 324  
 Db 415 enurertanentltspaqptenrettanentltsveptenrettanentltspaepht 474  
 QY 325 -----PTTITSA-PTTTPKEBAPTTPKAPTTTPKEBAPTTPKEBAPT---PAPTTPK 367  
 Db 475 qtrpfanektltssaeptehgerptplanen--tlpspaephtenrettanentltspaepht 532  
 QY 368 EPAPTTTPKAPAPTTTPKEBAPT---TPKEBAPTTPKAPAPTTTPKEBAPTTPKETAAPT 422  
 Db 533 engdrplanektltspaeptengqtrpfanektltssaeptehgerptplanentltspaepht 589  
 QY 423 KKLPTTPKELAPTPPKAPAPTTPEELAPTTPEEPPTTPKEBAPTTPKAAAPNT---PK 479  
 Db 590 speeptenrettanentltspaqptenremtanentlfpaeptenrettanentltspaepht 649  
 QY 480 EPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAP 533  
 Db 650 eptengqtrpfanektltssaeptehgerptplanentltspaephtenrettanentltspaepht 709

QY 534 LAPTTKEPSTSTSDKAPAPTTPKCTAPTTTPKEBAPTTPKKEBAPT---TPKGTAPTTLK 588  
 Db 710 aeptenertanentltspaqptengqtrpfanektltspaeptengqtrpfanektltspaept 769  
 QY 589 EPAPT-----TPKKPA-PKE-----LAPTTKGPSTTSDKAP 621  
 Db 770 saeptehaertplanentltspaqptenrettanentltsqfpaephtenrettanentltsqf 829  
 QY 622 TTPKET-----APTTKEBAPTTPKAPAPTTTPKEBAPTTPKKEBAPTTPKKEBAPT 670  
 Db 830 aeptenertanentltspaqptehaertplanentltspaqptengqtrpfanektltspaept 889  
 QY 671 HKSPDE-----STPELSAEPPTPK---ALENSPKPEGVPPT-----TKTPA 706  
 Db 890 saeptehgerptplanentltsraepthengerlanekaltspaeptengqtrpfanektltspaept 948  
 QY 707 ATPD-----EMTTAKDVTTERDLKTPPETTAAPKMTKE---TATTEKYTE 751  
 Db 949 saeptengertplanentltsaepthengerlanekaltspaeptengqtrpfanektltspaept 1008  
 QY 752 SKITATP-TQVSTTTQDTPPKI-----TTL-----KTTTLAPK 785  
 Db 1009 spakptehaertplanentltspaqptengqtrpfanektltspaeptengqtrpfanektltspaept 1068  
 QY 786 VTTT-----KKITTEIEMKPEETAKPKDRATNSKATTPPKOKP----- 825  
 Db 1069 itpslajphtegertltspndkltssaeptehaertplanentltspaqptengqtrpfanektltspaept 1128  
 QY 826 -----TKADKPKPTSTKPKKPTMPKPKP-----KTTPTPKKMTSTMBEL 863  
 Db 1129 ekmtqvtelstehtepkeltstte-klrtpekltysekltcltkgkntpvekpelenlgn 1187  
 QY 864 NPTSRINAMLOTTPRPNOT 883  
 Db 1188 tlletlikavkstenpekt 1207

RESULT 11  
 AAW31852  
 ID AAW31852 standard; Protein; 763 AA.  
 XX  
 AC AAW31852;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis 74 kda protein.  
 XX  
 KW Tuberculosis; mycobacteria; infection; diagnosis;  
 KM antimycobacterial; antibiotic; vaccine.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN W09741252-A2.  
 XX  
 PD 06-NOV-1997.  
 XX  
 PF 18-APR-1997; 97WO-EP01973.  
 XX  
 PK 29-APR-1996; 96DE-4017184.  
 XX  
 PA (GBF) GBF GES BIOTECH FORSCHUNG GMBH.  
 XX  
 PI Espitia C, Honisch C, Moreno C, Singh M;  
 XX  
 DR WPI: 1997-549750/50.  
 DR N-PSDB; AAT93610.  
 XX  
 PT New DNA and related proteins or RNA derived from M. tuberculosis -  
 PT used for diagnosis of mycobacterial infections, monitoring  
 PT vaccination and development of anti-mycobacterial agents  
 XX  
 PS Claim 5; Fig 13; 55pp; English.

```

FT      Domain      409..565
FT      /note="cohesin type II domain"
FT      Domain      607..763
FT      /note="cohesin type II domain"
XX
XX      FR2748479-A1.
XX
XX      14-NOV-1997.
XX
XX      10-MAY-1996; 96FR-0005854.
XX
XX      10-MAY-1996; 96FR-0005854.
XX
XX      (INSP ) INST PASTEUR.
XX
XX      Beguin P, Leibovitz E;
XX
XX      WPI: 1998-011569/02.
XX      N-PSDB: AAT86623.
XX
XX      Cellulase proteins with cohesin or dockerin type II domains - useful
XX      for potentiating the activity of multiprotein enzyme complexes
XX
XX      Claim 7, page 31-39; 60pp: French.
XX
XX      Multimeric protein, especially enzymatic, complexes are held together
XX      by protein-protein interactions between domains designated dockerins
XX      and cohesins, which are found on the catalytic and scaffold subunits
XX      respectively. An example of such a complex is the cellulose degrading
XX      protein complex from Clostridium thermocellum, known as the cellulosome.
XX      This complex comprises around 15 proteins including endoglucanases,
XX      cellobiohydrolases, hemicellulases, e.g. xylnases or lichenases, which
XX      interact with a central "scaffold" protein designated the cellulosome
XX      interacting protein (CIPA; see AAM43108). The catalytic subunits
XX      interact with the CIP subunit via conserved 23 amino acid dockerin
XX      domains. CIP has been shown to contain 9 copies of a cohesin domain.
XX      The invention relates to the isolation of proteins binding to a novel
XX      dockerin type domain found in the C-terminal portion of CIP. The new
XX      domain is designated a type II dockerin domain (as compared to the type
XX      I domain found on the catalytic subunits of the cellulosome). The type
XX      II dockerin domain has some sequence similarity to the type I dockerins
XX      but is unable to bind type I cohesin domains.
XX      The sequence presented here is an example of a protein which binds
XX      the novel type II dockerin domain and is the product of the O1P gene.
XX      The protein contains 4 type II cohesin domains in the N-terminal portion
XX      of which the first domain (amino acid residues 28-192) is thought to
XX      bind CIPA. The novel type II dockerin and cohesin domains can be used
XX      in complexes, especially enzyme complexes, to potentiate their catalytic
XX      actions in a synergistic manner.
XX
XX      Sequence 1664 AA:
XX
XX      Query Match      18.3%; Score 942; DB 19; Length 1664;
XX      Best Local Similarity 33.0%; Pred. No. 3e-45;
XX      Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;
XX
XX      QY      128 VLAKPPR-KAETTTGPAITTPKEPTTPKEPASTTPKEPTTTKSAPTTPKEPAPTT 186
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      758 VYIQPAPILKAADepIltlpsdeps-----depuss-----depIltlpsde 804
XX
XX      QY      187 TKSAPTPKEPAPTTTPKEPA-----PTTTKEPAPTTTSAPITTPKEPAPTT 242
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      805 ePTpsctpeepIltlpsdepsdepsdepsdepsdepsdepsdepsctpeepIltlps 864
XX
XX      QY      243 KAPATTPKEPAPTTTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 301
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      865 adeptpsdeplps--deplps--deplps--deplpsdeplpsdeplpsdeplps 919
XX
XX      QY      302 TTPKKEPA--PTTTKEPSPTTTPKEPAPTTTPKS-APTTPKEPAPTTTPKSAPTTPKEPSPTT 357
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      920 sdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 968

```

```

QY      358 TKEPAPTTPKEPAPTTTPKKAAPTTPKEPAPTTTPKEPAPTTTPKKAAPTTPKKE 416
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      969 sdeplps--deplps-----deplpsdeplpsdeplpsdeplpsdeplpsdeplpsd--- 1019
XX
XX      QY      417 TAPTPPKKLTPTPEKLAPTTPKEPAPTTPEELAPTTPEEPPTTP--PEEPAPPTTPKKAAP 475
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1020 -----eplpsde--eplpsdeplpsd-----eplpsdeplpsdeplpsdeplps 1064
XX
XX      QY      476 NTPKEPAPTTPKEPA--PTTPKKEPAAPT--TPKEPAPTTPKGAAPTTPKKAAPTTPKKAAP 532
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1065 tpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1122
XX
XX      QY      533 ELAPTTTKEPT-STTSKKAAPTTPKGAAPTTPKEPAPTTTPKEPAPTTPKGAAPTTPKKEPA 591
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1123 d-eplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1175
XX
XX      QY      592 PT-TPKKPAPEKELAPTTTPKGAAPTSTTSKKAAPT--TPKKEPAPTTPKEPAAPT--TPKKPAAPT 646
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1176 psetlpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1229
XX
XX      QY      647 TPETTPPTTSEVSTPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 706
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1230 dtpsdeplpsd--eplpsdeplpsd-----psdeplps--sdeplps--sdeplps--lpltdpsd 1277
XX
XX      QY      707 ATKPEMTTPAKDKTTERBDLRTTPETTTAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 766
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1278 eplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1337
XX
XX      QY      767 QDTTPPKKITTLLKTTTLAPKVTTPKKTITTTTINKKPEETAKPKDRATNSKATPKPO-KP 825
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1338 deplpsdeplpsd-----psdeplpsdeplpsdeplpsdeplpsdeplpsdeplps 1372
XX
XX      QY      826 TKAPKKEPTSKKKTTPKPRVKKKTTPPKKMTS-----TWPELNPSTRIA 870
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1373 setlpsdeplpsd-----lpslpslpslpslpslpslpslpslpslpslpslpslps 1424
XX
XX      QY      871 EAMLQTTTPNCPNPKSLVEVNPKSSEDAGAEGETTHMLLRPVHEVETPPMDIYLP 927
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      Db      1425 -----kptslpslpslpslpslpslpslpslpslpslpslpslpslpslpslpslpslps 1465
XX
XX      RESULT 10
XX      AAM03645
XX      ID      AAM03645 standard; Protein: 1325 AA.
XX      XX
XX      AC      AAM03645;
XX      XX
XX      DE      09-OCT-2001 (first entry)
XX      XX
XX      DE      Peptide #2327 encoded by probe for measuring breast gene expression.
XX      XX
XX      KW      Probe: human; breast disease; breast cancer; development disorder;
XX      KW      inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX      XX
XX      OS      Homo sapiens.
XX      XX
XX      PN      WO200157270-A2.
XX      XX
XX      PD      09-AUG-2001.
XX      XX
XX      PF      29-JAN-2001; 2001WO-US00661.
XX      XX
XX      PR      04-FEB-2000; 2000US-0180312.
XX      PR      26-MAY-2000; 2000US-0207456.
XX      PR      30-JUN-2000; 2000US-0608408.
XX      PR      03-AUG-2000; 2000US-0632366.
XX      PR      21-SEP-2000; 2000US-0234687.
XX      PR      27-SEP-2000; 2000US-0236359.
XX      PR      04-OCT-2000; 2000GB-0024263.
XX      XX
XX      PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX      XX
XX      PI      Penn SC, Hanzel DK, Chen W, Rank DR;

```

PR	28-JUN-1999;	9905-014595.1
PR	02-AUG-1999;	9905-014638.6
PR	02-AUG-1999;	9905-014638.8
PR	02-AUG-1999;	9905-014638.9
PR	03-AUG-1999;	9905-014703.8
PR	04-AUG-1999;	9905-014720.4
PR	04-AUG-1999;	9905-014730.2
PR	05-AUG-1999;	9905-014719.2
PR	05-AUG-1999;	9905-014726.0
PR	06-AUG-1999;	9905-014730.3
PR	06-AUG-1999;	9905-014741.6
PR	09-AUG-1999;	9905-014748.5
PR	09-AUG-1999;	9905-014793.5
PR	10-AUG-1999;	9905-014817.1
PR	11-AUG-1999;	9905-014831.9
PR	12-AUG-1999;	9905-014834.1
PR	13-AUG-1999;	9905-014855.5
PR	13-AUG-1999;	9905-014868.8
PR	16-AUG-1999;	9905-014936.8
PR	17-AUG-1999;	9905-014941.5
PR	18-AUG-1999;	9905-014942.6
PR	20-AUG-1999;	9905-014972.2
PR	20-AUG-1999;	9905-014972.3
PR	20-AUG-1999;	9905-014992.9
PR	23-AUG-1999;	9905-014990.2
PR	23-AUG-1999;	9905-014993.0
PR	25-AUG-1999;	9905-015056.6
PR	26-AUG-1999;	9905-015088.4
PR	27-AUG-1999;	9905-015106.5
PR	27-AUG-1999;	9905-015106.6
PR	27-AUG-1999;	9905-015108.0
PR	27-AUG-1999;	9905-015130.3
PR	31-AUG-1999;	9905-015143.8
PR	01-SEP-1999;	9905-015193.0
PR	07-SEP-1999;	9905-015236.3
PR	10-SEP-1999;	9905-015307.0
PR	13-SEP-1999;	9905-015375.8
PR	15-SEP-1999;	9905-015401.8
PR	16-SEP-1999;	9905-015403.9
PR	20-SEP-1999;	9905-015477.9
PR	22-SEP-1999;	9905-015513.9
PR	23-SEP-1999;	9905-015548.6
PR	24-SEP-1999;	9905-015565.9
PR	28-SEP-1999;	9905-015645.8
PR	29-SEP-1999;	9905-015659.6
PR	04-OCT-1999;	9905-015711.7
PR	05-OCT-1999;	9905-015717.3
PR	06-OCT-1999;	9905-015865.5
PR	07-OCT-1999;	9905-015860.29
PR	08-OCT-1999;	9905-015836.2
PR	12-OCT-1999;	9905-015833.9
PR	13-OCT-1999;	9905-015929.3
PR	13-OCT-1999;	9905-015929.4
PR	14-OCT-1999;	9905-015932.9
PR	14-OCT-1999;	9905-015933.29
PR	14-OCT-1999;	9905-015953.30
PR	14-OCT-1999;	9905-015953.31
PR	14-OCT-1999;	9905-015963.7
PR	14-OCT-1999;	9905-015963.8
PR	18-OCT-1999;	9905-015958.4
PR	21-OCT-1999;	9905-016074.1
PR	21-OCT-1999;	9905-016076.68
PR	21-OCT-1999;	9905-016076.67
PR	21-OCT-1999;	9905-016077.0
PR	21-OCT-1999;	9905-016081.5
PR	21-OCT-1999;	9905-016081.4
PR	22-OCT-1999;	9905-016098.1
PR	22-OCT-1999;	9905-016098.9
PR	25-OCT-1999;	9905-016140.4
PR	25-OCT-1999;	9905-016140.6
PR	26-OCT-1999;	9905-016135.9

PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match	19.0%;	Score 981;	DB 21;	length 763;
Best Local Similarity	38.5%;	Pred. No. 8.9e-48;		
Matches 226; Conservative	24;	Mismatches 293;	Indels 44;	Gaps 11;

[illegible]

## RESULT 9

ID AAW43106 standard; Protein; 1664 AA.

AC AAW43106;

DT 16-OCT-1998 (first entry)

DE C. thermocellum OJpB protein.

AA Multimer; enzyme; complex; protein-protein interaction; dockerin domain  
 KW cohesin domain; catalytic subunit; scaffold subunit; SdBA; synergistic;  
 KW cellulosome integrating protein; scaffoldin dockerin binding protein.

*Clostridium thermocellum*.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	Domain	28..192
----	--------	---------

FT	Domain	207..363
----	--------	----------

```

FT
/note= "cohesin type II domain"

```

PR	19-MAY-1999	9905-0134941
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136592
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138840
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139152
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999: 99JTS-01394651

PR 18-JUN-1999: 99UTS-0139463

PR 18-JUN-1999: 99JIS-0139763

PR	22-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.

23-JUN-1999; 99US-0140354

PR	28-JUN-1999;	99US-0140823.
PR	24-JUN-1999;	99US-0140695.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0142154.

PR 06-JUL-1999; 99US-0142390.

PR 09-JUL-1999; 99US-0142920.

PR 13-JUL-1999; 99US-0143542.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145218.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145918.

DB 157 vknkknrkkktpkpyvndeagsgldngdfvttctctcqhknvstskpittakpin 216  
 QY 61 PRPSLPNNDTSKETSLLTVNKKETVETKETTNNKOTSDGKKETTSKAKETOSIEETSAK 120  
 DB 217 prpslpnndtsketskelslvnketvketcttnkqtsldgketsketsketsketsk 276  
 QY 121 DIAPTSKVLAKPTPKAETTTKGPALTPPKETPTTTPKEPASTPKPPPTTIKASAPTTTK 180  
 DB 277 diaptskvlakptpkaketttkgpalttkpripctpkpkepasttkpcttktsapttk 336  
 QY 181 EPAPTTTSKAPPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKASAPTTTK 240  
 DB 337 epaptttskapttkpkeappttkpkeappttkpkeappttktskpnlpptscxxxx 396  
 QY 241 PKKPAPTTKKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPKPAPTTPKE 298  
 DB 397 ctcp-----tpkcpnpiprslmpktkcpapctpkpkeapctapckpapiplpe 443

RESULT 7  
 AAM24516  
 ID AAM24516 standard; Protein; 5179 AA.  
 AC AAM24516;  
 XX  
 DT 12-OCT-2001 (first entry)  
 DE C899P predicted amino acid sequence.  
 XX  
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 XX immunogenic; gene therapy; vaccine; colonic cancer.  
 OS Homo sapiens.  
 PN WO200149716-A2.  
 PD 12-JUL-2001.  
 XX  
 PF 29-DEC-2000; 2000WO-US35596.  
 XX  
 PR 30-DEC-1999; 9905-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Xu J., Lodges MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jlang Y;  
 DR WPI; 2001-441847/47.  
 XX  
 PT Colon tumor associated proteins and nucleic acids useful for the  
 PT prevention, diagnosis and treatment of colonic cancer -  
 XX  
 PS Claim 2; Page 446-462; 472p; English.

CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512  
 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences  
 CC given in the exemplification of the present invention.  
 XX  
 SQ Sequence 5179 AA:

Query Match 22.7%; Score 1168; DB 22; Length 5179;  
 Best Local Similarity 32.5%; Pred. No. 1.8e-57;  
 Matches 363; Conservative 53; Mismatches 448; Indels 254; Gaps 40;

QY 21 DEAGSGLDNGDFK-----VTPDTSYTOH-NKVSNSPK----- 52  
 DB 1312 dhpsgsdgdreipfgvcgapediecrsvkdpnlslsegqgkvqcdvsyglcknedgf 1371  
 QY 53 -----ITWAKPINDRPSLPPNSDTSKETSLLTVNKKETTYETK 88  
 DB 1372 gngpfqglydyklrvncwpmckcltppstltpspptlttltpstltpspptlt 1427  
 QY 89 ETTTNNKOTSDGKKETTSKAKETOSIEKTSKDLAPSKVLAPTKAETTTGAPALTTP 148  
 DB 1428 -tttppstltpspptlttltp-----lpt-----tpspstltpspptlt 1470  
 QY 149 KEPT-----PTPKEPASTPKETPTTIKASAP-TTPKEP-----APTTSKAP-TTP 194  
 DB 1471 spptltpspptltpspptltpptltpspptltpspptltpspptltpspptltp 1530  
 QY 195 KEPAPTTKKEPAPTTPKEP-----APTTSKAP-TTPKEPAPTTPKKA 245  
 DB 1531 ttpptltpspptltpstltpstltpstltpstltpstltpstltpstltpsppt 1590  
 QY 246 PTPKKEPAPTTPKEPPTTPPKKEPAPTTPKKEP-----APTAKKAPAPTTPKE 298  
 DB 1591 itlttppstltpspptltpstltpstltpstltpstltpstltpstltpstltp 1650  
 QY 299 PAPTTPKEPAPTTPKEPPTTPPKKEPAPTTPKASAPTTPKKEPAPTTPKAP--TTPKEPSP 355  
 DB 1651 ppptltpspptltpspptltpstltpstltpspptltpspptltpstltpstltp 1709  
 QY 356 TTPKKEPAPTTPKEPAPTTPKKAAPTTPKKEPAPTTPKKAAPTTPKKAAPTTPK 414  
 DB 1710 ttttstltpspptltpstltpstltpstltpstltpstltpstltpstltpstl 1769  
 QY 415 -----KETAPTTP-----KTLTP 427  
 DB 1770 fspstltpstltpcplcnwtglwldsgkpnfhkpggdteilgdcpgpmaaniscratmvp 1829  
 QY 428 -----TTPKLAAPT 437  
 DB 1830 dvplgdlqgvcdvsyglcknedqkpgyvmalfnelynvqceccvtqplmtltt 1889  
 QY 438 PEKPAPTTPPEELAPTTPEEPPT-TPEEPAPTTPKKAAPTTPKKEPAPTTPKEP-----AP 491  
 DB 1890 tempptltpstltpstltpstltpstltpstltpstltpstltpstltpstltpst 1949  
 QY 492 TTPKEPAPT-TPKETAATTTKGTAPTLKBPAPPTPKKPAKELAPTTKEPTSDKP 550  
 DB 1950 tltpvtpstltpstltpstltpstltpstltpstltpstltpstltpstltpstltp 2009  
 QY 551 AP-TTPKGTAPTTPKKEPAPTTPKKEPAPTTPKGAATTLKBPAPPTPKKPAKELAPTTTK 609  
 DB 2010 tpttptltpstltpstltpstltpstltpstltpstltpstltpstltpstltpst 2068  
 QY 610 GPTSTSDKAP-TTPKETAPTTPKKEPAPTTPKKAAPTTPKEPTTPPTTSEVSTPTTKEPT 668

CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonectin may be used in gene therapy. The present sequence represents  
CC a substantial portion of a human MSF-derived tribonectin.

XX Sequence 902 AA:

Query Match 57.4%; Score 2956.5; DB 22; Length 902;  
Best Local Similarity 72.2%; Pred. No. 9,9e-159;  
Matches 699; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

QY 1 VKDKKKRRTKKKPPKPPVVDAGSGLDNGDFKVTTPDSTGTHNKVSTSPKTTAKPIN 60  
DB 1 vkdkknrtkkpckppvvdagsgldngdfkvttpdstgthnkvsstpkttakpin 60  
QY 61 PRSLPNSDTSKETS/LTVNKKETVETKETTNTKOTSGKKEKTKAKENOSTEKTSK 120  
DB 61 prslpnsdstsketsltvnkktvetkettntkotskgtkktaketsak 120  
QY 121 DLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKEPASTPKPEPTTTKSAPTTPK 180  
DB 121 dlaptskvlakptpkaetttkcpal-----tlpkepsttpkepttkksapttpk 172  
QY 181 EPAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTT 240  
DB 173 epapttksaptpkepap-tlkepaplt-kepap-tlkepapltkepaplt-kepaplt 228  
QY 241 PKRPAPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTPKPEPA 300  
DB 229 -kepaplt-kepaplt-kepaplt-kepaplt-kepaplt-kepaplt-kepa 281  
QY 301 PTPPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKEPAPTTKE 360  
DB 282 ptt-kepap-tlkepaplt-kepapltkepap-tlkepapltkepaplt-kepap-tlke 335  
QY 361 PAPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTPKPEPAPTTTPKEPAPTT 420  
DB 336 paptt-kepaplt-kepaplt-kepaplt-kepap-tlkepaplt-kepaplt-kepap 388  
QY 421 TPKKLPTTPTEKLAPTTPKEPAPTTBELAPTTPPEPTTPTEPAPTTKAAAPNTPK 480  
DB 389 t-kepapltkepap-tlkepaplt-kepaplt-kepaplt-kepaplt-ke 440  
QY 481 PAPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAPPTPKPEPAPTTTPKEPAPTT 540  
DB 441 paptt-kepaplt-kepaplt-kepaplt-kepaplt-kepaplt-kepap-----tlk 488  
QY 541 EPTSTSDKAPPTPKGTAPTTPKEPAPTTTPKEPAPTTTPKGAPTTPKKEPAPTTPKKAP 600  
DB 489 epapttkkepaplt-kepaplt-kepaplt-kepaplt-kepaplt-kepaplt-kepap 541  
QY 601 KELAPTTTKGPTSTSDKAPPTPKETAAPTTPKEPAPTTTPKAPPTTPKEPAPTTSEPT 660  
DB 542 -----tlkepapltkepaplt-kepaplt-kepaplt-kepapltkepapltkepa- 589  
QY 661 PTTTKEPTTHKSPDSTPRLSAEPTPKALENSPKRGVTTTPAPATKEMTTAKDKT 720  
DB 590 -pttkepapltkepaplt-----tlkepapltkepapltkepapltke- 636  
QY 721 TERDLRTPEPTTAAKPKMETATTEKTESKITAATTVOSTTTODTTPFKITTLTKT 780  
DB 637 -----paptt-----tlkepaplt-----kepapltkepapltkepap----- 667  
QY 781 TLAPKATTTTKTTTTEINMKKPEETAKPKDRATNSKATTPKPKAPPKPTSTKPKPT 840  
DB 668 -----tlkepapltkepap-----tlkepapltkepap-apc-tkepap- 686  
QY 841 MPKVRKPKTTTPPKKMTSTWPELNPSTRIAEMLOTTTRNOPNSKIVNVNKSSEDAGS 900  
DB 687 -----aptpckrmtstwpeelnpsriaeaml-tctipngpnsklvevnpkseadag 739  
QY 901 AEGETPHMLRPVHFPEVTPDMVLPVRVNOGIIINPMISDETINCGNPKVGLTTLRN 960  
DB 901 aegetphmlrpvhfpevtpdmvlpvrvnogiiinpmisdetinngnkvglttlarn 960

DB 740 aegtpmhllrpvhfpevtpdmvlpvrpnqgllnplmsdetinngnkvglttlarn 799  
QY 961 GTLVAFRG 968  
DB 800 gtlvafrg 807

RESULT 6  
ID AAR80041 standard; Protein; 452 AA.  
XX AAR80041;  
AC AAR80041;  
XX 10-Apr-1996 (first entry)  
DT 10-Apr-1996 (first entry)  
XX Human megakaryocytopoietin protein.  
DE Human megakaryocytopoietin; wheat germ agglutinin; heparin;  
XX megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;  
KW multipotential stem cell.  
KM Homo sapiens.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc-difference 393..396  
FT /note="unspecified amino acids"  
FT Misc-difference 444..446  
FT /note="unspecified amino acids"  
PN WO9523861-A1.  
XX 08-SEP-1995.  
PD 08-SEP-1995.  
XX 06-MAR-1995; 95MO-CN00015.  
PE 06-MAR-1995; 94CN-0112066.  
XX 04-MAR-1994; 94CN-0112066.  
PR (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.  
PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.  
XX Gu X, Han Z, Shen Q;  
PI WPI; 1995-320576/41.  
XX N-PSDB; AAT04546.  
DR New haematopoietic cell growth factor - used for treating  
XX thrombocytopenia and hematocytopenia  
XX Example; Page 23; 36pp; Chinese.  
PS This sequence represents the human megakaryocytopoietin (MPO) protein.  
CC This sequence was purified using a carrier which can couple wheat germ  
CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
CC AAR80039 and AAR80040) were used to produce the amplification primers  
CC shown in AAT04544 and AAT04545. The fragments amplified by these primers  
CC can then be used as probes to screen human cDNA libraries for MPO cDNA.  
CC The MPO cDNA can then be inserted into a plasmid which is used to  
CC transform cells to produce MPO. The MPO sequence is capable of promoting  
CC colony formation of megakaryocytes, enlarging the size of megakaryocytes  
CC and stimulating the proliferation of multipotential stem cells. The  
CC factor may be used for treating thrombocytopenia and hematocytopenia.  
CC The purification method can be used to isolate MPO from human urine or  
CC serums of patients with aplastic anaemia, and from animal blood or urine  
CC by radiation exposing the animals to induce aplastic anaemia.  
XX Sequence 452 AA:

Query Match 26.3%; Score 1357.5; DB 16; Length 452;  
Best Local Similarity 88.3%; Pred. No. 3.7e-69;  
Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;  
QY 1 VKDKKKRRTKKKPPKPPVVDAGSGLDNGDFKVTTPDSTGTHNKVSTSPKTTAKPIN 60  
DB 1 vkdkknrtkkpckppvvdagsgldngdfkvttpdstgthnkvsstpkttakpin 60

CC measuring the amount of MSF or its fragment in a biological sample of a  
 CC mammal, wherein an increased amount of MSF compared to a control  
 CC indicates the presence of or predisposition to developing  
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
 CC treatment of osteoarthritis, where they may be used for lubricating  
 CC mammalian joints, such as articulating joints of humans, dogs or horses.  
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
 CC useful for inhibiting adhesion between two surfaces such as the injured  
 CC tissues of a mammal, where the injury is caused by a surgical insertion  
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
 CC tribonectin may be used in gene therapy. The present sequence represents  
 CC human MSF.  
 XX  
 XQ Sequence 1404 AA;

Query Match	100.0%	Score 5155;	DB 22;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 4.2e-282;		
Matches 968; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy	1	VKDNKKNRRTKKKPPRKPVPYDDEAGSLDNDKFVYTPDSTYQHNHNVSNRSPKTTAKPI	60
Db	200	Vkdnhknrtkkkpkpdpvdeaagsg,ldngdfkvtcpdtsltqnmvsspsklttakp	259
Qy	61	PRPSLPKNSDSKESLTVNKEVVEIKETTTNNKQSTGDKKRTYSAKETOSIEKTSAK	120
Db	260	prpslpnpsdsksktsltvnhkettvekkettlnkqstgdgkksctsaektsaektsak	319
Qy	121	DLAPTSKVLAKPTPKAETTTKGALJTTPKKEPTPTPKESASTPKKEPTTTKSAPTTK	180
Db	320	dlaptskvlakpdpkaeettgpaaltcpkeptptcpkepstcpkeptltksapttk	379
Qy	181	EPAPRTTKSAPTTKKEAPPTTTKKEAPPTTKKEAPPTTKKEAPTTTKKSAPTKKEAPPT	240
Db	380	epaprttksapttckepaptttkkeapttckepaptttkkeaptttkksapttckepaptt	439
Qy	241	PKKPAPPTPKKPAPPTPKKEPTPTPKKPAPPTPKAPPTTKKEBAPTTAKKPAPPTTKKEB	300
Db	440	pkkpapptckepapttckeppttpckepapttckepapttckepapttckepapttckepa	499
Qy	301	PTTPKEAPPTTTKEBSPTTKKEAPPTTTKSAPTTKKAPPTTTKSAPTTPKESPTTKE	360
Db	500	pttpkeaptttkespslttkkeaptttksapttkkapptttksapttpkespsttk	559
Qy	361	PAPTPKKEBAPPTTKKAPPTPKBAPPTPKBAPPTTKKBPAPAPKEBAPPTPKETAP	420
Db	560	paptpkkebaptttkkapttkkbpaptttkkbpaptttkkbpapabkeapttcketapt	619
Qy	421	TPKLTPTTPPEKLAPTTPPEKAPPTPEELAPTTPPEEPTPTTPPEBAPTTPKAAPTPK	480
Db	620	tpkltpttppeklapttppekapttpeelapttpeeptpttpeebapttpkaaaptpk	679
Qy	481	PAPPTPKBAPPTTKKEBAPTTPKETAPTTKEGAPTTLKBPAPPTTKKAPKELAPTTK	540
Db	680	paptpkbeaptttkebpaptttketapttkgapttlkbpaptttkkappakelaapttk	739
Qy	541	PPTSTTSUKBAPPTPKGAPPTPKBAPPTPKBAPPTPKGTAAPTTLKBAPPTPKKBP	600
Db	740	epsttsckpapttcpkgaptcpkeaptcpkeaptcpgaptcpgaptclckepapttckpab	799
Qy	601	KELAPTTTKGAPTSTSDKAPPTTKBAPTTPKBAPPTTKKAPPTPKBAPPTTSBVT	660
Db	800	kelaptttkgaptstsdkaptttkbaptttkbaptttkkapttkkapttpecppttsvst	859
Qy	661	PTTTKEPTTTIHKSPDESPELSABPTPALENSPKBEGVPTTKTPAATKPEMTTTAKDKT	720
Db	860	ptttkeptttihkspdespeelsaepcpalenspekbgvpttktpaatkpeemttlakdk	919
Qy	721	TFRDLRTPEPTTAPAKMTKEATTTBKTMTSKTTATTTQVTSSTMODDTPKLTTLKTT	780
Db	920	terdlrttpepttapaakmtkeatttketsklattlqvtstttdttcpklttlktt	979

Qy	781	TLAKRVTTTKKTTITTEIIMAKKEETIAKPKDRATINSATTPKQKRTPKAKPKPTSTTKKKT	840
Db	980	tlakrvtttkklltlltelnmkpeetlaxpkdranskatlpkqkpkakpkpkslkkpct	10399
Qy	841	MPVRKPKCTTPTPEKMTSTMBELNPTSRIAEALQVTTTPRPNQTPPSKSLVEVNPKSEDAAG	900
Db	1040	mpvrkpkcttptpekrmtstmbelnpsrilaeanlqctttrpnlqtpnksklvevnpkseadag	10999
Qy	901	AEGETPMHLARPHVEFPEVTPDDMDLPRVNPQGIITNPMISDETNI CNCKRPVYDGLTTLRN	960
Db	1100	aegetpmlarphvfpvtpddmdylprvpngiitnplmsdetnlcnckrpvydglttlrn	1159
Qy	961	GTLYAERG 968	
Db	1160	gtlyavrg 1167	

RESULT	5
AAB29778	
ID	AAB29778 standard; Protein; 902 AA

AC AAB29778;

DT 28-FEB-2001 (first entry)

Human MSF-derived tribonectin.

Human tribonectin; MSF; megakaryocyte stimulating factor;  
KM alternative splicing; joint boundary lubricant; O-linked oligosaccharide

KM osteoarthritis; tribosupplementation; tissue adhesion inhibitor  
KM friction coefficient reduction; gene therapy; antiarthritic;;  
KM osteopathic.

OS Homo sapiens.

PN W0200064930-A2.

PD 02-NOV-2000.

24-APR-2000; 2000WO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PI Jay GD;

DR WPI; 2001-024673/03.

Novel tribonectin polypeptide useful as lubricant for treating

PS Disclosure; Fig 1; 47pp; English.

The invention relates to a human tribonectin which is a product of alternative splicing of the human MSF (megakaryocyte stimulating factor) gene. The tribonectin has at least one O-linked oligosaccharide substituting moiety and has a polypeptide sequence comprising 1-76 repeats of a motif having at least 50% identity to the sequence KEPAPPT (AAB29774). The invention also relates to a nucleic acid encoding a human MSF-derived tribonectin: a biocompatible composition comprising a human tribonectin for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical incision or trauma, or an artificial device e.g., an orthopaedic implant. In



XX Example 1: Page -: 34pp: English.  
 XX  
 CC Administration of a composition comprising the campodactylly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anaesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactylly-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U70316).  
 CC  
 XX  
 XX Sequence 1404 AA:

Query Match 100.0%; Score 5155; DB 22: Length 1404;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-282;  
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKKKRKKKKPPKPPVVDAGSGLDNGDEKVTTPDSTGTHNKVSTSPKITTAKPIN 60  
 DB 200 VKDKKKRKKKKPPKPPVVDAGSGLDNGDEKVTTPDSTGTHNKVSTSPKITTAKPIN 259  
 QY 61 PRPSLPNSDTSKSTSLTVNKKETVETKETTNTKOSTGCKKKTSAKKEOSTEKRSK 120  
 DB 260 PRPSLPNSDTSKSTSLTVNKKETVETKETTNTKOSTGCKKKTSAKKEOSTEKRSK 319  
 QY 121 DLAPTSTVLAKPPPKAETTTGAPALTTTPKEPPTTPKEPASTTPKEPPTTTPK 180  
 DB 320 DLAPTSTVLAKPPPKAETTTGAPALTTTPKEPPTTPKEPASTTPKEPPTTTPK 379  
 QY 181 EPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 240  
 DB 380 EPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 439  
 QY 241 PKRPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 300  
 DB 440 PKRPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 499  
 QY 301 PTPPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 360  
 DB 500 PTPPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 559  
 QY 361 PAPPTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 420  
 DB 560 PAPPTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 619  
 QY 421 TPKKLTPTTEKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 480  
 DB 620 TPKKLTPTTEKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 679  
 QY 481 PAPPTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 540  
 DB 680 PAPPTPKAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 739  
 QY 541 EPSTSTSDKAPPTTPKTAATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 600  
 DB 740 EPSTSTSDKAPPTTPKTAATTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 799  
 QY 601 KELAPPTTKGPTSTSDKAPPTTPKETAATTPKEPAPTTPKEPAPTTPKEPAPTTP 660  
 DB 800 KELAPPTTKGPTSTSDKAPPTTPKETAATTPKEPAPTTPKEPAPTTPKEPAPTTP 859

QY 661 PTTTKEPTTIHKSPDESTPELSAEPTRKALENSREKPGVPTTKTPATKPEMTTANDKT 720  
 DB 860 PTTTKEPTTIHKSPDESTPELSAEPTRKALENSREKPGVPTTKTPATKPEMTTANDKT 919  
 QY 721 TERLRTPTTEPTTAPKMTKETATTTKTESKITTATTOVTSSTTODTPPEKITTAKT 780  
 DB 920 TERLRTPTTEPTTAPKMTKETATTTKTESKITTATTOVTSSTTODTPPEKITTAKT 979  
 QY 781 TLAPKVTTKTKTTTTEIMNKPETANPKDRATNSKATTPKPKPTKPKPTSTKKPKT 840  
 DB 980 TLAPKVTTKTKTTTTEIMNKPETANPKDRATNSKATTPKPKPTKPKPTSTKKPKT 1039  
 QY 841 MPVRKKTTPPTPKMTSTWELNPTSRIRAFAMLQTTTPRQOTNSKLVENPKSEDAAG 900  
 DB 1040 MPVRKKTTPPTPKMTSTWELNPTSRIRAFAMLQTTTPRQOTNSKLVENPKSEDAAG 1099  
 QY 901 AEGTTPHMLRPHVMEVETPDMDYLPRVPRNOGIIPMLSDFTNICKGRPVDDLTLRN 960  
 DB 1100 AEGTTPHMLRPHVMEVETPDMDYLPRVPRNOGIIPMLSDFTNICKGRPVDDLTLRN 1159  
 QY 961 GTLVAFNG 968  
 DB 1160 GTLVAFNG 1167

RESULT 4  
 AAB29773  
 ID AAB29773 standard; Protein: 1404 AA.  
 XX  
 AC AAB29773;  
 XX  
 DT 28-FEB-2001 (first entry)  
 XX  
 DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.  
 XX  
 KW Human MSF; megakaryocyte stimulating factor; tribonectin;  
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
 KW friction coefficient reduction; gene therapy; antiarthritic;  
 KW osteopathic.  
 XX  
 XX Homo sapiens.  
 OS  
 PN MO200064930-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 24-APR-2000; 2000MO-US10953.  
 XX  
 PR 23-APR-1999; 99US-0298970.  
 XX  
 PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
 XX  
 XX Jay GD;  
 PI  
 DR WPI: 2001-024673/03.  
 XX  
 DR N-FSDB: AAC81498.  
 XX  
 PT Novel tribonectin polypeptide useful as lubricant for treating  
 PT osteoarthritis, comprises O-linked lubricating moiety  
 XX  
 PS Claim 3; Page 7; 47pp: English.  
 XX  
 CC The invention relates to a human tribonectin which is a product of  
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
 CC gene. The tribonectin has at least one O-linked oligosaccharide  
 CC lubricating moiety and has a polypeptide sequence comprising 1-76  
 CC repeats of a motif having at least 50% identity to the sequence KEPAPT  
 CC (AAB29774). The invention also relates to a nucleic acid encoding a  
 CC human MSF-derived tribonectin; a biocompatible composition comprising a  
 CC human tribonectin for inhibiting tissue adhesion formation; and a method  
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by

DR WPI: 1992-284660/34.  
DR N-PSDB: AA027223.

XX New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
PS bacterial and viral infections, etc.

Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine preceding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNA's each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, i.e. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organisation of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterised by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.

Sequence 1404 AA:

Query Match 100.0%; Score 5155; DB 13; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 4.2e-282;  
Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRRTKKRPKPPVVDGASGLNDGDFVTTPTDSTTHNKKVSTSPKITTAKPIN 60  
DB 200 vkdknkrrtkrkprpvvdeagsgldngdfvttptdstthnkvstspkittakpin 259  
QY 61 PRPSLPPNSDTSKETSILYVNEKETTVEETKETTNNKQTSIDGKEKTSAKTOSTIEKTSAK 120  
DB 260 prpslppnsdtsketsiltvneketvettettnkqtsidgketsktsaktsak 319  
QY 121 DLAPTGSVLAKPPPKAETTKGPAITPKPEPTTPPEKPASTPKPEPTTISAPPTPK 180  
DB 320 dlaptgsvlakpppkatettkgpaitpkpepttppekpaastpkpepttlksaptpk 379  
QY 181 EPAPTTTSAPTPKPEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTSAPTPK 240  
DB 380 epaptttsaptpkpeaptttkepaptttkepaptttkepaptttkepaptttsaptpk 439  
QY 241 PKRPAPTTKEPAPTTKEPTPTTPPEKPAETTKKEPAPTTTKEPAPTTTKEPAPTTKEPA 300  
DB 440 pkrapapttkepapttkeptpttppekpaettkkepaptttkepaptttkepapttkepa 499  
QY 301 PPTPKPEAPTTKEPSPPTPKPEAPTTTTSAPTTTKEPAPTTTTSAPTTKEPSPPTTKE 360  
DB 500 ptpkpeapttkepsptpkpeapttttsaptttkepapttttsapttkepspttke 559  
QY 361 PAPPTPEKAPATTPKKAPATTPKKAPATTPKKAPATTTKKAPATAPKPEAPTTPEKTAPT 420  
DB 560 papptpekapattpkkapattpkkapattpkkapattpkkapattpkkapattpkxetapt 619  
QY 421 TPKKLTPTTEKLAAPTPEKPAETTPPEELAPTTPEEPTPTTPPEAPATTTKAAAPNPK 480  
DB 620 tpkkltptteklaptppekpaettppeelapttpeeptpttppeapatttkaaapnpk 679

QY 481 PAPPTPKPEAPTTPKKAPATTPKKETAAPTTPKGTAPTTTLKEPAPTTTPKKAPKELAPTTTK 540  
DB 680 papptpkpeaptpkkapattpkketaptpkgtaptttlkepapttpkkapkelaptttk 739  
QY 541 EPTSTSDKPAATTPKGTAPTTTPKEPAPTTTPKEPAPTTTGTAPTTLKEPAPTTTPKKAP 600  
DB 740 eptstsdkapattpkgtaptptpkepaptttpkepapattpkgtapttlkepapttpkkap 799  
QY 601 KELAPTTTGTSTSDKPAATTPKETAAPTTPKEPAPTTTPKKAPATTPEPTPTTSEVST 660  
DB 800 kelaptttgtstsdkapattpketaptpkepaptttpkkapattpetppllsevast 859  
QY 661 PTTTKEPTTIHKSPDSTPELSAEPPTKALENSKEPGVPTTPATTPPEMTTAKDPT 720  
DB 860 ptttkepttihkspdstpeelsaepptkalenskepgvpttpattppemttakdpt 919  
QY 721 TERDLRTPEETTAAPKMTKETATTEKTESKITATTTQVSTTTQDTTPFKITTLKTT 780  
DB 920 terdlrtpeettaapkmtketattekteskitatttqvtstttqdttpfkittlkt 979  
QY 781 TLAPKVTTKTITTTTBIIMNKKPEETAKPKDRATNSKATTPPKOKPKARKKPTSTKKPT 840  
DB 980 tlapkvttktittttbiimnkkpeetakpkdratnskattppkokpkarkkptstkkpt 1039  
QY 841 MPKVRKPKTTPPRKMTSTMPSELNPTSRILAEAMLOTTTRPNQSPNSKIVEVNPKSEDAG 900  
DB 1040 mpkvrkpkttpprkmtstmpselnptsrilaemlootttrpnqspnslvevnpskedsag 1099  
QY 901 AEGETPMLLRPHVMEPTVPMDDYLPRVPGNGIINPMLSDETINCNKGPVDGLTTLRN 960  
DB 1100 aegetpmlrrphvmeptvpmdyldrpgngiinpmlsdetincnkgpvdglttlrn 1159  
QY 961 GTLVAFRG 968  
DB 1160 gtlvafrg 1167

RESULT 3  
AAB60568  
ID AAB60568 standard; protein; 1404 AA.  
XX  
AC AAB60568;  
DT 27-APR-2001 (first entry)  
XX  
DE Human megakaryocyte stimulating factor (MSF, CACP).  
KW Human; CACP protein; camploclactyly-arthropathy-coxa vara-pericarditis;  
KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
KW antiarthritic.  
XX  
OS Homo sapiens.  
XX  
PN WO200107068-A1.  
XX  
PD 01-FEB-2001.  
XX  
PF 21-JUL-2000; 2000WO-US20002.  
XX  
PR 23-JUL-1999; 9905-0145328.  
PR 19-JUL-2000; 2000US-0145328.  
PA (UYCA-) UNIV CASE WESTERN RESERVE.  
XX  
PI Warman MI;  
XX  
DR WPI: 2001-182721/18.  
XX  
PT New composition comprising the camploclactyly-arthropathy-coxa  
PT vara-pericarditis protein in combination with an anesthetic, useful for  
PT treating osteoarthritis, or as lubricants of tissue and joints -

DR N-PSDB: AAH98981.  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
PS Claim 20: Page 1198-1201; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
XX Sequence 1299 AA:

Query Match 100.0%; Score 5155; DB 22; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 3,9e-282;  
Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNTKKKPPKPPVVDAGSGLDNGDFKYTPPTSTTOHKKVSTSKITAKPIN 60  
DB 200 VKDNKKNTKKKPPKPPVVDAGSGLDNGDFKYTPPTSTTOHKKVSTSKITAKPIN 259  
QY 61 PRPSLPNSDTSKESLTVNKEETVETKETTNTKOTSDGKEKTSKESOSTEKTSK 120  
DB 260 PRPSLPNSDTSKESLTVNKEETVETKETTNTKOTSDGKEKTSKESOSTEKTSK 319  
QY 121 DLAPTQVLAAPTPKAETTTGPAITTPKEPTTPKEPASTTPKEPTTPKSAPTTPK 180  
DB 320 DLAPTQVLAAPTPKAETTTGPAITTPKEPTTPKEPASTTPKEPTTPKSAPTTPK 379  
QY 181 EPAPTTTSAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 240  
DB 380 EPAPTTTSAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 439  
QY 241 PKKPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPA 300  
DB 440 PKKPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPA 499  
QY 301 PTPKKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKE 360  
DB 500 PTPKKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKE 559  
QY 361 PAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 420  
DB 560 PAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 619  
QY 421 TPKKLPTTTEKLAAPTPPEKPAATTPPEELAPTPPEEPPTTPPEEPADTPKAA 480  
DB 620 TPKKLPTTTEKLAAPTPPEKPAATTPPEELAPTPPEEPPTTPPEEPADTPKAA 679  
QY 481 PAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 540  
DB 680 PAPTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 739  
QY 541 EPSTSTSDKAPTPPKGTAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 600  
DB 740 EPSTSTSDKAPTPPKGTAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 799  
QY 601 KELAPTTTGGPTSTTSKAPTPPKETAPTPPKGTAATTPKEPAATTPKEPAATTP 660  
DB 800 KELAPTTTGGPTSTTSKAPTPPKETAPTPPKGTAATTPKEPAATTPKEPAATTP 859  
QY 661 PTTTKEPTTTHKSPDESTPELSAEPTPKALENSPKREGVTTTTPATKREMTTKDK 720  
DB 860 PTTTKEPTTTHKSPDESTPELSAEPTPKALENSPKREGVTTTTPATKREMTTKDK 919  
QY 721 TEPDLRTPTPTTTAAAPKMKETATTEKTESKITATTTQVSTTQDTTPFKITLKT 780  
DB 721 TEPDLRTPTPTTTAAAPKMKETATTEKTESKITATTTQVSTTQDTTPFKITLKT 780

DB 920 terdlrtpettlaapkmktelatttektlesklatattgvtsttqtdtpfkltlkt 979  
QY 781 TLAPKVTTKKRTTTTTTEINMKPEETAKPKDRATNSKATTPKPKPKPKPKPKPK 840  
DB 980 TLAPKVTTKKRTTTTTTEINMKPEETAKPKDRATNSKATTPKPKPKPKPKPKPK 1039  
QY 841 MPKRAKPKTTTPPKKMTSTPELNPTSRIDAMLOTTTRPNQTPNSKLVEVNPKSEDAG 900  
DB 1040 MPKRAKPKTTTPPKKMTSTPELNPTSRIDAMLOTTTRPNQTPNSKLVEVNPKSEDAG 1099  
QY 901 AEGEPHMLLRPHVEMPEVTPDMYLPKVPNOGIIINPMISDENINCKRVDGLTTRN 960  
DB 1100 AEGEPHMLLRPHVEMPEVTPDMYLPKVPNOGIIINPMISDENINCKRVDGLTTRN 1159  
QY 961 GTLVAFRG 968  
DB 1160 gtlvafrg 1167

## RESULT 2

AAR26049  
ID AAR26049 standard; Protein: 1404 AA.

XX AAR26049;

DT 02-FEB-1993 (first entry)

XX MSF precursor.

XX Megakaryocyte colony stimulating factor: secretion signal; meg-CSF;  
KW stability; proteolytic cleavage; adhesion; alternative splicing.

XX Synthetic.

XX Key Location/Qualifiers

FT Region 1..26

FT /label= Exon\_I

FT Region 26..67

FT /label= Exon\_II

FT Region 67..107

FT /label= Exon\_III

FT Region 107..157

FT /label= Exon\_IV

FT Region 157..200

FT /label= Exon\_V

FT Region 200..1141

FT /label= Exon\_VI

FT Region 1411..1166

FT /label= Exon\_VII

FT Region 1166..1212

FT /label= Exon\_VIII

FT Region 1213..1266

FT /label= Exon\_IX

FT Region 1266..1331

FT /label= Exon\_X

FT Region 1331..1373

FT /label= Exon\_XI

FT Region 1373..1404

FT /label= Exon\_XII

PN WO9213075-A.

PD 06-AUG-1992.

XX 17-JAN-1992; 92WO-US00433.

XX 18-JAN-1991; 91US-0643502.

XX 10-SEP-1991; 91US-0757022.

XX (GENW ) GENETICS INST INC.

XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

XX

XX

XX

XX

XX

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:03 ; Search time 107.17 Seconds  
(without alignments)  
669.058 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1167

Perfect score: 5135  
Sequence: 1 VKDKKKRKKKKPKPPVV.....GKPVDTLTTLNGLTVAFRG 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5155	100.0	1299	22	AAAM24322 Human EST encoded
2	5155	100.0	1404	13	AAAR26048 MSF precursor. SY
3	5155	100.0	1404	22	AAAB60568 Human megakaryocyt
4	5155	100.0	1404	22	AAAB29773 Human megakaryocyt
5	2956.5	57.4	902	22	AAAB29778 Human MSF-derived
6	1357.5	26.3	452	16	AAAR80041 Human megakaryocyt
7	1168	22.7	5179	22	AAAM24516 C899P predicted am
8	981	19.0	763	21	AAAG38942 Arabidopsis thalia
9	942	18.3	1664	19	AAAM3106 C. thermocellum OI
10	771.5	15.0	1325	22	AAAM03645 Peptide #2327 enco
11	715	13.9	763	18	AAAM31852 Mycobacterium tube

12	625.5	12.1	4412	21	AAV53666	Sequence g1/101742
13	612	11.9	572	12	AAAM31855	Mycobacterium tube
14	578.5	11.2	472	22	AAAB60569	Bovine MSF ortholo
15	544	10.6	844	7	AAAB60570	Sequence of the Fa
16	542	10.5	807	21	AAV54467	Aminl acid sequenc
17	520.5	10.1	788	21	AAV54466	Amino acid sequenc
18	506.5	9.8	744	9	AAAB2975	Bioadhesive precu
19	504	9.8	1837	21	AAAB11726	Cryptosporidium pa
20	496.5	9.6	2971	21	AAAB1231	Human ORFX ORF95
21	489.5	9.5	2972	22	AAAB50363	Human SRCAP. Homo
22	489.5	9.5	3118	22	AAAB50362	Human SRCAP. Homo
23	489	9.5	1721	21	AAAB1727	Portion of cryptos
24	488.5	9.5	826	13	AAAR26042	P. yoelii SSP2 ant
25	488	9.5	617	22	AAAM16458	Peptide #2892 enco
26	488	9.5	617	22	AAAM04187	Peptide #2869 enco
27	488	9.5	957	21	AAV59288	Human MucII polype
28	488	9.5	957	22	AAAM24513	C900P predicted am
29	488	9.5	1721	19	AAAM48299	Cryptosporidium pa
30	485	9.4	1127	22	AAAB95541	Human protein sequ
31	476.5	9.2	652	9	AAAB82974	Bioadhesive precu
32	467.5	9.1	511	22	AAAM14883	Peptide #1317 enco
33	467.5	9.1	511	22	AAAM27312	Peptide #1349 enco
34	467.5	9.1	511	22	AAAM02607	Peptide #1289 enco
35	450.5	8.7	378	12	AAAR14160	PRP encoded by clo
36	446.5	8.7	378	12	AAAR14162	C. albicans Rbt1 p
37	446.5	8.7	750	20	AAV05477	Caenorhabditis ele
38	443	8.6	2870	21	AAV95559	Caenorhabditis ele
39	443	8.6	3178	21	AAV95556	Japanese sea musse
40	442.5	8.6	751	16	AAAR80839	S. epidermidis ope
41	425.5	8.3	910	22	AAAG83007	Human atrophin-1 r
42	424.5	8.2	1012	20	AAV17406	Mouse microtubule-
43	424.5	8.2	1125	21	AAAB22934	Microtubule associ
44	424.5	8.2	1125	21	AAV79637	EYFP-DEVD-MAP4-BB
45	424.5	8.2	1610	21	AAAB2870	

## ALIGNMENTS

RESULT 1  
AAM24322 standard; Protein; 1299 AA.  
ID AAM24322:  
AC AAM24322:  
XX  
XX  
DE 12-OCT-2001 (first entry)  
DE Human EST encoded protein SEQ ID NO: 1847.  
XX  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200154477-A2.  
XX  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
XX  
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX  
DR WPI: 2001-476164/51.













PT para-pericarditis protein in combination with an anesthetic, useful for  
PT treating osteoarthritis, or as lubricants of tissue and joints  
XX  
XX  
XX Example 1; Fig 4; 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the  
CC administration of a composition comprising the camptodactyl-arthropathy-  
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
CC The composition may further comprise a local anaesthetic. The composition  
CC of the invention may be administered via intra-articular or intravenous  
CC injection. The human CACP protein is identified in the invention as  
CC being megakaryocyte stimulating factor (MSF). The gene encoding  
CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
CC this gene are responsible for the heritable disorder camptodactyly  
CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
CC acts as a synovium lubricant, and can be used to lubricate tissue and  
CC joints in the treatment of osteoarthritis. The composition may be  
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
CC loss of range of movement or joint damage). The present sequence  
CC represents the bovine orthologue of human MSF, superficial zone  
CC protein (SZP).  
XX  
XX

Sequence 472 AA:

Query Match 15.2%; Score 823.5; DB 22; Length 472;  
Best Local Similarity 74.9%; Pred. No. 2.5e-39;  
Matches 155; Conservative 17; Mismatches 32; Indels 3; Gaps 3;

QY 808 PKDRATNSKATPTKPKKAPKPKSTKPKPTMPVRKPKTPPRK-MTSTMPRLNPT 866  
DB 76 pgratnsavtctpkpkrpkpkpkstkpkprt-prvtprktpktpktttampept 134  
QY 867 SRIAEAMLQTTTRPNQTPNSKLVENVPKSEDAAGAGETPRHMLLRHVFMEVTPDMYL 926  
DB 135 s-lpreamlqtttrpnpseidvnsenedgdaaegekphmalfprpvltprvptell 193  
QY 927 PRVPPNGIILINPLSDENINCKNPYDGLTTLRNGTLVAFRGHYFWMLSPESSPARRT 986  
DB 194 vcpssqgfiglmpfidelncgripvdgltllngllvalfrgnylmllpctpppprrt 253  
QY 987 TEVWGISPSIDVFTFRCNCEGKTFEKK 1013  
DB 254 tevvgispsidvftfrcncegkttfkk 280  
RESULT 11  
AAM03645  
ID AAM03645 standard; Protein; 1325 AA.  
XX  
XX AAM03645;  
XX  
XX 09-OCT-2001 (first entry)  
XX Peptide #3227 encoded by probe for measuring breast gene expression.  
XX  
XX Probe; human; breast disease; breast cancer; development disorder;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
XX Homo sapiens.  
XX  
XX WO200157270-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 29-JAN-2001; 2001WO-US00661.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -

XX Claim 27; SEQ ID No 12385; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes  
CC (see AAI00010-AI10067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1325 AA:

Query Match 14.2%; Score 771.5; DB 22; Length 1325;  
Best Local Similarity 26.0%; Pred. No. 6e-36;  
Matches 302; Conservative 97; Mismatches 466; Indels 295; Gaps 45;

QY 7 NNRK-KKPPPKP-----PVYDEAGSGIDNDRKVTTPDSTNQ-----HNKYSTSKITNT 55  
DB 60 ntrhskpdkpdkpdknskldhksstcdhnapptseensngqkdpmlrnqsvdpadct 118  
QY 56 A-----KPINRPSILPPN---SDTSKET-----SLTVNKEETVETKET 90  
DB 119 tthkesagkhhitpnpkskincrkstkystvtrksdkgrpleksmsltdktsushkt 178  
QY 91 TTT-----NKQSTDEKETTSAKET-----QSIEKT-SAKDLAPTSKYLAKT 133  
DB 179 ttfhngsqnqtkqsktsifekitaasktyktqgpeesektedrtvaadkllktl 238  
QY 134 PKAETTTKGPALT-----TPKEPPT--TPKEPASTTPKEPPT 170  
DB 239 knlgetlsaneltqslaepthngtraneennpsspepenterententttspepten 298  
QY 171 TIKSA--PTTPKEPAPT-----TIKSAPTTPKEPAPTTPKE--PAPTTPKEPAPT 216  
DB 299 rertanentapfpaqplenremtanentllfpaepthngtraneentttspeptenger 358  
QY 217 TTKEPAPTTPKSAPTTPKEPAPTTPKKAAPTTPKKEPAPTTPKEPAPT 271  
DB 359 tane-----tltpspaeptengerlfpandkltsssaesteherlplanentttspeapt 414  
QY 272 -TKEPAPTTPKEPAPTAPKKAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 324  
DB 415 enterententttspsagprentretlanektlisvpeptenretlanektlispsaepteng 474  
QY 325 -----PTTTKSA--PTTTKEPAPTTPKSAPTTPKEPAPTTPKE-----PAPTTPK 367  
DB 475 grtplanekttssaeptengerlplanen--tltpspaeptenretlanektlispsaept 532  
QY 368 EPAPTTPKKAAPTTPKEPAPT-----TPKEPAPTTPKKAAPTAPKKAAPTTPKEPAPTTP 422  
DB 533 engdrtplanektlispsaeptengerlfpandkltsssaesteherlplanentttspeapt 589  
QY 423 KKLTPPTPEKLAPTTPKEKAPAPTTPPEELAPPTTPPEPPTTPPEPAPTTPKKAAPT--PK 479

FT Domain 409..565  
 FT /note="cohesin type II domain"  
 FT 607..763  
 FT /note="cohesin type II domain"  
 XX FR2748479-A1.  
 XX 14-NOV-1997.  
 PD 10-MAY-1996; 96FR-0005854.  
 XX 10-MAY-1996; 96FR-0005854.  
 XX 10-MAY-1996; 96FR-0005854.  
 XX (INSP ) INST PASTEUR.  
 PA Beguin P, Leibovitz E;  
 P1 WPI: 1998-011569/02.  
 DR N-PSDB; AAT86623.  
 XX  
 PT Cellulase proteins with cohesin or dockerin type II domains - useful  
 PT for potentiating the activity of multiprotein enzyme complexes  
 XX  
 PS Claim 7; Page 31-39; 60pp; French.

CC Multimeric protein, especially enzymatic, complexes are held together  
 CC by protein-protein interactions between domains designated dockerins  
 CC and cohesins, which are found on the catalytic and scaffold subunits  
 CC respectively. An example of such a complex is the cellulose degrading  
 CC protein complex from Clostridium thermocellum, known as the cellulosome.  
 CC This complex comprises around 15 proteins including endoglucanases,  
 CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which  
 CC interact with a central "scaffold" protein designated the cellulosome  
 CC integrating protein (CIP; see AAM43108). The catalytic subunits  
 CC interact with the CIP subunit via conserved 23 amino acid dockerin  
 CC domains. CIP has been shown to contain 9 copies of a cohesin domain.  
 CC The invention relates to the isolation of proteins binding to a novel  
 CC dockerin type domain found in the C-terminal portion of CIP. The new  
 CC domain is designated a type II dockerin domain (as compared to the type  
 CC I domain found on the catalytic subunits of the cellulosome). The type  
 CC II dockerin domain has some sequence similarity to the type I dockerins  
 CC but is unable to bind type I cohesin domains.  
 CC The sequence presented here is an example of a protein which binds  
 CC the novel type II dockerin domain and is the product of the oipb gene.  
 CC The protein contains 4 type II cohesin domains in the N-terminal portion  
 CC of which the first domain (amino acid residues 28-192) is thought to  
 CC bind CIPa. The novel type II dockerin and cohesin domains can be used  
 CC in complexes, especially enzyme complexes, to potentiate their catalytic  
 CC actions in a synergistic manner.

XX Sequence 1664 AA;

Query Match 17.4%; Score 942; DB 19; Length 1664;  
 Best Local Similarity 33.0%; Pred. No. 1.8e-45;  
 Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

QY 128 VLAKPTP-KAETTKGPAUTTKKEPTTKKEPASTTKKEPTTKSAPTTKKEAPPT 186  
 Db 758 VVIGPAPKIAAASDEPLDTPSDEPLPS-----DEPLPS-----DEPLPSDEPLPSD 804  
 QY 187 TKAPPTPKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTK 242  
 Db 805 EPTSPSETPPEEPLDTPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSD 864  
 QY 243 KPAPTPKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAP 301  
 Db 865 SDEPLPSDEPLPS--DEPLPS--DEPLPS--DEPLPSDEPLPSDEPLPSDEPLPSDEPL 919  
 QY 302 TTPKKEPA--PTTKKEPSP--TKKEPAPPTTKKS--APTTKKEPAPPTTKKSAPTTPKKEPSP 357  
 Db 920 SDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPS 968

QY 358 TKKEPAPPTPKKEPAPPTTKKKEPAPPTTKKEPAPPTTKKKEPAPPTTKKEPAPPTTKKE 416  
 Db 969 SDEPLPS--DEPLPS--DEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSD 1019  
 QY 417 TAPPTPKKLTPTPEKLAAPPTPKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTK 475  
 Db 1020 -----EPLPSDE--PLPSDEPLPSD-----EPLPSDEPLPSDEPLPSDEPLPSD 1064  
 QY 476 NTPKEPAPPTTKKEPA--PTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKE 532  
 Db 1065 TPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPS 1122  
 QY 533 ELAPTTKEPT-STTSNDKPAAPPTTKGTAAPPTTKKKEPAPPTTKKEPAPPTTKKEPAP 591  
 Db 1123 d-EPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSD 1175  
 QY 592 PT-TPKKPAKKEAPPTTKGPTSTSDKPAPT--TKKEPAPPTTKKEPAPPTTKKEPAPPT 646  
 Db 1176 PSETPPEE-----LPTDPSDEPLPSDEPLPSDEPLPSD-EPLPSDEPLPSDEPLPSD 1229  
 QY 647 TPERPPTTSEVSTPTTKKEPTTIHKSPDSTPELSAEPKALENSPKKEGVPPTTKTPA 706  
 Db 1230 DTPSDEPLPSD--EPLPSDEPT---PSDEPT--SDEPT--SDEPT--LPTDPSD 1277  
 QY 707 ATKPEMTTAKDKTTERDLPPTTETTAAPKMKETATTEKTESKITATTQVSTTT 766  
 Db 1278 EPTSPSEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSDEPLPSD 1337  
 QY 767 QDTPTFKITTKTTLAKVVTTKKITTTEIMNKPEETAKPKDRATNSKATPKPKP-KP 825  
 Db 1338 DEPLPSDEPT-----PSDEPTDPSDEPLPSDEPLPSDEPT 1372  
 QY 826 TKAPKKTSTKTKMPRVKPKPTTPPKMTS-----TPELNPISRIA 870  
 Db 1373 SEPEEPLPTTP-----TPSPCLTPSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1424  
 QY 871 EAMLGTTTRPNQTPNSKLVEVNPKSEDPAGAGEETPHMLRPVHFVPEVTPDMDYLP 927  
 Db 1425 -----KPTSPAPTELE-EPLPSDVGALSGENRAYLRY-----PDGSIPT 1465

RESULT 10  
 AAB60569  
 ID AAB60569 standard; Protein; 472 AA.  
 XX  
 AC AAB60569;  
 XX  
 DT 27-APR-2001 (first entry)  
 XX  
 DE Bovine MSF orthologue, superficial zone protein (SZP).  
 XX  
 KW Bovine; GACP protein; campodactylly-arthropathy-coxa vara-pericarditis;  
 KM Superficial zone protein; SZP; MSF orthologue; synovial lubricant;  
 KW osteoarthritis; joint lubrication; osteopatinic; anilarthritis.  
 OS Bos taurus.  
 XX  
 PN WO200107068-A1.  
 PD 01-FEB-2001.  
 XX  
 PF 21-JUL-2000; 2000MO-US20002.  
 XX  
 PR 23-JUL-1999; 99US-0145328.  
 XX  
 PR 19-JUL-2000; 2000US-0145328.  
 XX  
 PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX  
 PI Warman ML;  
 XX  
 DR WPI: 2001-182721/18.  
 XX  
 PT New composition comprising the campodactylly-arthropathy-coxa

```

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

```

```

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 18.1%; Score 981; DB 21; Length 763;
Best Local Similarity 38.5%; Pred. No. 5,1e-48;
Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```

```

QY 137 ETTTGPAALTTPKEPTPTPKKEPASTTPKEPTTTKSAPTTPKBPATTTKSAPTTPKE 196
DB 126 qtdqgnrlylpppprlpppppcvtcp-epspp-----ppptlvt---pqilpt 170
QY 197 PAPTTPKEPAPTTPKEPAPTTPKEPAPT--TKSAPTTPKEPAPTTPKBPATTTKEPAP 254
DB 171 lptltpetpctpqtatpctlevlptcqtptlptqptlptlptlptlptlptlptlptl 226
QY 255 TTPKEPTTPPKKEPAPT--KEPAPTTPKEPAPTAKKBPATTPKBPAPTTPKEPAPTTPK 313
DB 227 nlppeptltpetcpntpupespnlppeiipntlppqtlpntlppetltpetltpetlpp 286
QY 314 EESPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPSPTTPKEPAPTTPKEPAPT 373
DB 287 etspnlppeptltpetltpetltpetltpetltpetltpetltpetltpetltpetl 340
QY 374 PKKBPATTPKKEPAPTTPKBPAPTTPKBPAPVAPKEPAPTTPKEPAPTTPKEPAPT 433
DB 341 ppqtlpntltpetltpetltpetltpetltpetltpetltpetltpetltpetltpet 400
QY 434 APPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAAAPTTPKEPAPTTPKEPAPT 493
DB 401 lptltpqtlpntltpetltpetltpetltpetltpetltpetltpetltpetltpet 460
QY 494 PKKBPATTPKKEPAPTTPKBPAPTTPKBPAPVAPKEPAPTTPKEPAPTTPKEPAPT 553
DB 461 ptkcp-ptltpetltpetltpetltpetltpetltpetltpetltpetltpetltpet 514
QY 554 TPKGTAPPTTPKEPAPTTPKEPAPTTPKGTAPPTTPKEPAPTTPKBPAPVAPKEPAPT 613
DB 515 lppqtlpntltpetltpetltpetltpetltpetltpetltpetltpetltpetltpet 574
QY 614 TTSDEP-----APTTPKEPAPTTPKEPAPTTPKBPAPTTPKEPAPTTPKEPAPTTP 666
DB 575 lppntlppqtlpntltpetltpetltpetltpetltpetltpetltpetltpetltpet 634
QY 667 P---TTTHKSPDSTPELSAETPKALENSPKKEGVPVTTTPAATKP 710
DB 635 ppptltpssptspnpspppkspqppppprftqppppprgtcpc 681

RESULT 9
AAW43106
ID AAW43106 standard; Protein; 1664 AA.
XX AAW43106;
AC
XX
XX
DT 16-OCT-1998 (first entry)
DE
XX
XX C. thermocellum OLPB protein.
XX
XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
XX cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
XX cellulosome integrating protein; scaffoldin dockerin binding protein.
OS
XX Clostridium thermocellum.
XX
XX
XX Key Location/Qualifiers
XX 28..192
XX /note="cohesin type II domain"
XX FT 207..363
XX FT Domain /note="cohesin type II domain"

```

PR	19-MAY-1999	9905-0134841
PR	20-MAY-1999	9905-0135124
PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136592
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139453
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456

## AAG38942

ID	AAG38942	standard; Protein; 763 AA.
XY		

AC AAG38942;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 48115.

protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000;

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132407.  
 01-MAY-1999; 99US-0132407.  
 01-MAY-1999; 99US-0132407.

PR	05-MAY-1999;	99US-0132485.
BB	06-MAY-1999	0000-0133400

07-MAY-1999: 00UTC-01322487.

PK	99US-0134256.
II-MAY-1999;	
TA-MAY-1999:	99US-0134318
PB	

PK	14-MAY-1999;	9905-0134219.
PR	14-MAY-1999;	9905-0134331.

14-MAY-1999; 99US-0134370.  
18-MAY-1999; 99US-0134768

90/4570-5055



CC particular, one of the surfaces is pericardial tissue: DNA encoding a  
 CC tribonectin may be used in gene therapy. The present sequence represents  
 CC a substantial portion of a human MSF-derived tribonectin.  
 XX

Sequence 902 AA:

Query Match 59.4%; Score 3217.5; DB 22; Length 902;  
 Best Local Similarity 73.4%; Pred. No. 27e-174;  
 Matches 744; Conservative 28; Mismatches 80; Indels 161; Gaps 65;

```

QY 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDKVTTPTDSTTOHKNKSTSPKITTAKPIN 60
DB 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDKVTTPTDSTTOHKNKSTSPKITTAKPIN 60
QY 61 PRSLPNSDTSKETSILTVNKEETVETETTTNKTSDGKEKTSKAKTOSIKTSK 120
DB 61 PRSLPNSDTSKETSILTVNKEETVETETTTNKTSDGKEKTSKAKTOSIKTSK 120
QY 121 DLAPTSKVLAKEPTPKAETTTKGPALTTPKEPPTTPKEPASTTPKEPTTKASPTTPK 180
DB 121 DLAPTSKVLAKEPTPKAETTTKGPALTTPKEPPTTPKEPASTTPKEPTTKASPTTPK 180
QY 181 EPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKE 240
DB 173 EPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKE 240
QY 241 PKKRAPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 300
DB 229 -keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-ke 281
QY 301 PTPPEAPPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 360
DB 282 ptt-keap-keap-keap-keap-keap-keap-keap-keap-keap-keap-keap 335
QY 361 PAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPAPAPAPAPAPAP 420
DB 336 paplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keap 388
QY 421 TPKLTPPTPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPE 480
DB 389 t-keapltkeap-apttke-apttke-apttke-apttke-apttke-apttke-apttke-ke 440
QY 481 PAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPK 540
DB 441 paplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keap 488
QY 541 EPTSTSDKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTP 600
DB 489 epapltke-apttke-apttke-apttke-apttke-apttke-apttke-apttke-keap 541
QY 601 KEAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAP 660
DB 542 -ttkeapltke-apttke-apttke-apttke-apttke-apttke-apttke-apttke- 589
QY 661 PTTTKEPTTHKSPDSTPELSAEPKPALENSPKKAPAPTPPKAPAPTPPKAPAPTP 720
DB 590 -pttkeapltkeaplt-keap-keap-keap-keap-keap-keap-keap-keap-ke 636
QY 721 TERDLTPPTTPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAP 780
DB 637 -keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keaplt-keap 667
QY 781 TLAPKVTYTKKTTTTEINMKPEETAKPKDRATNSKATTPKQKQPKAPAPTPPKAP 840
DB 668 -ttkeapltkeap-apt-keap-apt-keap-apt-keap-apt-keap-apt-keap 686
QY 841 MPKVRKAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAP 900
DB 687 -apttkeapltkeapltkeapltkeapltkeapltkeapltkeapltkeapltkeap 739
QY 901 AEGETPHMLLRHVMPPEVTPMDYLRVVPNGIITINMLSDETICGKGVDTGLTILRN 960
  
```

DB 740 aegetphmlrrhvfmpvtpmdylrvvpngiitinpmlsdetlcnkgkpvdtlilrn 799  
 QY 961 GTLVAFRGHYFMWLSPPSPSPARRITTEWGISPIDVTPTFCNCEGKTFEKK 1013  
 DB 800 gltvalfgyhfmwlsppsparritlewgijspidvtlvtfcncgkttffk 852

RESULT 6  
 AAR80041  
 ID AAR80041 standard; Protein; 452 AA.  
 AC AAR80041:  
 XX  
 DT 10-APR-1996 (first entry)  
 DE Human megakaryocytopoietin protein.  
 KW Human: megakaryocytopoietin; wheat germ agglutinin; heparin;  
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;  
 KW multipotential stem cell.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 393..396 /note= "unspecified amino acids"  
 FT Misc-difference 444..446 /note= "unspecified amino acids"  
 FT  
 PN W09523861-A1.  
 PD 08-SEP-1995.  
 XX  
 PF 06-MAR-1995; 93WO-CN00015.  
 XX  
 PR 04-MAR-1994; 94CN-0112066.  
 XX  
 PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.  
 XX  
 PI Gu X, Han Z, Shen Q;  
 XX  
 DR WPT. 1995-320576/41.  
 DR N-PSDB; AAT04546.  
 XX  
 PT New haematopoietic cell growth factor - used for treating  
 PT thrombocytopenia and hematocytopenia  
 PS  
 PS Example: Page 23; 36pp; Chinese.  
 XX  
 CC This sequence represents the human megakaryocytopoietin (MPO) protein.  
 CC This sequence was purified using a carrier which can couple wheat germ  
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
 CC AAR80039 and AAR80040) were used to produce the amplification primers  
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers  
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.  
 CC The MPO cDNA can then be inserted into a plasmid which is used to  
 CC transform cells to produce MPO. The MPO sequence is capable of promoting  
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes  
 CC and stimulating the proliferation of multipotential stem cells. The  
 CC factor may be used for treating thrombocytopenia and hematocytopenia.  
 CC The purification method can be used to isolate MPO from human urine or  
 CC sera of patients with aplastic anaemia, and from animal blood or urine  
 CC by radiation exposing the animals to induce aplastic anaemia.  
 XX  
 SQ Sequence 452 AA;

Query Match 25.1%; Score 1357.5; DB 16; Length 452;  
 Best Local Similarity 88.3%; Pred. No. 1.6e-69;  
 Matches 263; Conservative 3; Mismatches 21; Indels 11; Gaps 1;  
 QY 1 VKDNKKNRRTKKRPKPPVVDGSGLDNGDKVTTPTDSTTOHKNKSTSPKITTAKPIN 60  
 |||

measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is pericardial tissue. DNA encoding a tribonectin may be used in gene therapy. The present sequence represents human MSF.

Sequence 1404 AA:

Query Match 100.0%; Score 5416; DB 22; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 2.6e-298;  
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKRRTKKRPVVDAGSLDNGDFKVTTPDSTTQHNKVSPTKITTAKPIN 60  
DB 200 vkdnkrrtkrkrrpvyvdaegslndngdfkvttpdsttqhknkvsptkittakpin 259  
QY 61 PRPSLPNSDTSKETSITVKNKETTVEKETTNNKQSTDGKEKTTSAKETOSIEKTSK 120  
DB 260 prpslpnsdtsketsitvknkettvetkettnnkqstdgketttsaketosiekttsak 319  
QY 121 DLAPTSKVLAKPPPKAETTNGKALITTPKRPPTTPKEPASTTPKEPTPTTISAPTTPK 180  
DB 320 dlaptskvlnkpppkattngkalittpkrppttpkepasttpkepptpttisapttk 379  
QY 181 EPAPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTP 240  
DB 380 epapttksapttpkpapttpkepapttpkepapttpkepapttkpapttkpapttk 439  
QY 241 PKRPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKRPAPTTPKEPA 300  
DB 440 pkrappttpkepapttpkepapttpkepapttpkepapttkpapttkpapttkp 499  
QY 301 PPTTKREPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 360  
DB 500 ppttkrepapttpkepapttpkepapttkpapttkpapttkpapttkpapttk 559  
QY 361 PAPPTKREPAPTTPKEPAPTTPKEPAPTTPKRPAPTTPKEPAPTTPKEPAPT 420  
DB 560 papptkrepapttpkepapttpkepapttkpapttkpapttkpapttkpapttk 619  
QY 421 TPKKLTPTTPPEKLAATTPTEKAPPTTPEELAPTTPEEPPTTPEEPAPTTPKAA 480  
DB 620 tpkkltpttppeklaattppekapttpeelapttpeeppttpeepapttkpaa 679  
QY 481 PAPPTKREPAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAP 540  
DB 680 papptkrepapttpkepapttpkepapttkpapttkpapttkpapttkpapttk 739  
QY 541 EPTSTTSKDPAPTTPKGTAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKRPAPT 600  
DB 740 eptsttskdpapttpkgapttpekapttkpapttkpapttkpapttkpapttk 799  
QY 601 KELAPTTTGGPTSTSDKAPATTPKETAPTTPEKAPATTTPKRPAPTTPPEPT 660  
DB 800 kelaptttggptstsdkapattpekapttpekapttkpapttkpapttkpapttk 859  
QY 661 PPTTKPPTTIHKSDESPTELSAEPPTKALENSPEGVPPTTKTAPATPEMTTAKOKT 720  
DB 860 ppttkppttihsdesptelsaepptkalspckpvgvpttktpaatpemttaokt 919  
QY 721 TERDLATPTPTTTAAAPKAKTKETATTEKTTESKITATTTQVSTTQODTPPKIT 780  
DB 920 terdlatptptttaaapkaktetatttektteskittatctqgstttqdtlfpkitt 979

QY 781 TLAPKYTTTKTTTTEIMNKKPEETAKPKDRATNSKATTPKPOKPTKAPKPTSTKPK 840  
DB 980 tlapkytttktttteimnkkpeetakpkdratnskattpkpoakptkaptkptsttkpk 1039  
QY 841 MPVRKPKPTTPPRKMTSTPELNPPTSRIAEMALQTTTPRNOTPNSKLYEVPKSDAG 900  
DB 1040 mpvrkpkpttpprkmtstpelnpptsriaemalqtttprnotpnsklyevpkstdag 1099  
QY 901 AEGSTPMMLRPHVFMETVPDMDYLPRVNOGIITINPMLSDFTNCSGKPYDGLTTLN 960  
DB 1100 aegstpmmlrphvfmetyvpdmdylprvnogiitinpmlsdfntcsgkpydglttln 1159  
QY 961 GEIVAFGHYFMMLSPSPSPSPARRITEWGISPSIDVTFRNCBCKTFFEK 1013  
DB 1160 geivafghyfmmlspspspsparritewgispsidvtfrncbcktffek 1212

# RESULT 5

AAB29778 AAB29778 standard; Protein: 902 AA.

AAB29778: AAB29778:

28-FEB-2001 (first entry)

Human MSF-derived tribonectin.

Human tribonectin; MSF; megakaryocyte stimulating factor; alternative splicing; joint boundary lubricant; O-linked oligosaccharide; osteoarthritis; tribosupplementation; tissue adhesion inhibition; friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

Homo sapiens.

WO200064930-A2.

02-NOV-2000.

24-APR-2000; 2000WO-US10953.

23-APR-1999; 99US-0298970.

(RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

Jay GD;

WPI; 2001-024673/03.

Novel tribonectin polypeptide useful as lubricant for treating osteoarthritis, comprises O-linked lubricating moiety

Disclosure: Fig 1; 47pp; English.

The invention relates to a human tribonectin which is a product of alternative splicing of the human MSF (megakaryocyte stimulating factor) gene. The tribonectin has at least one O-linked oligosaccharide lubricating moiety and has a polypeptide sequence comprising 1-76 repeats of a motif having at least 50% identity to the sequence KEPAPTT (AAB29774). The invention also relates to a nucleic acid encoding a human MSF-derived tribonectin; a biocompatible composition comprising a human tribonectin for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In



XX Example 1: Page -: 34pp; English.  
 PS The invention relates to a method of treating osteoarthritis via the  
 XX administration of a composition comprising the campodactylly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1425-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactylly-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a Genbank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U70316).  
 XX  
 XX Sequence 1404 AA:

Query Match 100.0%; Score 5416; DB 22; Length 1404;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-298;  
 Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKNNKTKKPKPPVVDAGSLDNGDFKTPDSTYQNNKSTSKITTAAPIN 60  
 Db 200 vkdknnktrtkkpkppvvdagsgldngdfkvtcpdstlqnnkvsstkltakpin 259  
 QY 61 PRSLPNSDTSKSTSLTVNKEETVETKTTTNTKQSTGKKEKTAQSIEKTSK 120  
 Db 260 prslpnsdtskstsiltvnkettvetkettntkqstgkkttsaketsksak 319  
 QY 121 DLAFTSVLVKPPPKAETTTKGAALTPKKEPTPTPKKPASTPKKEPTPTTISAP 180  
 Db 320 dlaftsvlvakppkaetttkgaaltppkkeppttpkkpastsppkktpttk 379  
 QY 181 EPAFTTTSAPTPKKEPAPTTKKEPAPTTKKEPAPTTTSAFTTKKEPAPT 240  
 Db 380 epaftttsaptpkkepaptpkkepaptpkkepaptpkksafttkkapt 439  
 QY 241 PKKAPPTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEP 300  
 Db 440 pkkapptpkkepaptpkkepaptpkkepaptpkkepaptpkksaptpk 499  
 QY 301 PTTKKEPAPTTKKEPAPTTKKEPAPTTTSAFTTKKEPAPTTTKEPPTTKE 360  
 Db 500 pttkkepaptpkkepaptpkkepaptpkksafttkkapttkkapspttk 559  
 QY 361 PAPTTPKEPAPTTKKEPAPTTKKEPAPTTTKEPAPTTTKEPAPTTKKEPAPT 420  
 Db 560 papttpkepaptpkkepaptpkkepaptpkksaptpkksaptpkapt 619  
 QY 421 TPKKLTPTTPEKLAFTTPEKAPATTPEEPPTTPEEPAPTTTKEPAPTTKKE 480  
 Db 620 tpkkltpptpeklafttpekapttpeeppttpeepaptttkkapaapttk 679  
 QY 481 PAPTTPKEPAPTTKKEPAPTTKKEPAPTTTKEPAPTTTKEPAPTTKKEPAPT 540  
 Db 680 papttpkepaptpkkepaptpkkepaptpkksaptpkksaptpkapt 739  
 QY 541 EPTSTSDKAPATTPKGAPATTPKKEPAPTTKKEPAPTTTKEPAPTTKKEPAP 600  
 Db 740 eptstsdkapattpkgapattpkkepaptpkkepaptpkksaptpk 799  
 QY 601 KELAPTTTGGPTSTSDKAPATTPKKEPAPTTKKEPAPTTTKEPAPTTTKEPAPT 660  
 Db 800 kelaptttggptstsdkapattpkkepaptpkksaptpkksaptpk 859

QY 661 PTTKKEPTTTHKSDESTPELSAEPKALENSKEPGEVPTTKTAATKEMTTAKDKT 720  
 Db 860 pttkkepntthksdestpelsaepkalenskepgevpnttktaatkemttdk 919  
 QY 721 TERDLRTPEPTTAAAPMTKETATTEKTESKTTATTTQVSTTTODTTPFKITLKT 780  
 Db 920 terdlrtpeptttaaapmtketattekteskttatttqvstttodttpfk 979  
 QY 781 TLADRVTTTKTTTITTEIMNKPEETAKPKDRAATNSKATTPKPKTKPKSTTKPK 840  
 Db 980 tladvrttktttittteimnkpeetapkdratnskattpkpktpkstkpk 1039  
 QY 841 MPRRKKTTPTPPKKMTSTPELNPSTRIAEAMQTTTPRPNQPNKSLVEVNPKS 900  
 Db 1040 mprrkkttptpkkmtstpelnpstriaeamqtttprpnqpnkslvevnpk 1099  
 QY 901 AEGETPMALRPHVFMPEVTPDMQYLPRVFNQGIITIPMLSDENTCNCKPVDGLTLRN 960  
 Db 1100 aegetpmlrphvfmpevtpdmqylprvfnqgillipmlsdenlcnckpvdgl 1159  
 QY 961 GTLVAFRGHYFWMLSPESSPARRTTEVWGIPSPIDVFTTRCNCEKTPFFK 1013  
 Db 1160 gltvafrghyfwmlspessparrttevwgipspidvfttrcncegtkffk 1212

## RESULT 4

AAB29773  
 ID AAB29773 standard; Protein; 1404 AA.

AC AAB29773;  
 DT 28-FEB-2001 (first entry)  
 XX

DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

XX Human MSF; megakaryocyte stimulating factor; tribonecin;  
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
 KW friction coefficient reduction; gene therapy; antiarthritic;  
 KW osteopathic.

XX Homo sapiens.

XX WO200064930-A2.

XX 02-NOV-2000.

XX 24-APR-2000; 2000WO-US10953.

XX 23-APR-1999; 99US-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX Jay GD;

XX WPI; 2001-024673/03.

XX N-PSDB; AAC81498.

XX Novel tribonecin polypeptide useful as lubricant for treating

XX osteoarthritis, comprises O-linked lubricating moiety

XX Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonecin which is a product of  
 CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
 CC gene. The tribonecin has at least one O-linked oligosaccharide  
 CC lubricating moiety and has a polypeptide sequence comprising 1-76  
 CC repeats of a motif having at least 50% identity to the sequence KKPAPT  
 CC (AAB29774). The invention also relates to a nucleic acid encoding a  
 CC human MSF-derived tribonecin; a biocompatible composition comprising a  
 CC human tribonecin for inhibiting tissue adhesion formation; and a method  
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by

DR WPI: 1992-284660/34.  
N-PSDB: AA027223.

XX New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
PT bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1; 87Pp; English.

XX The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine preceding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNA's each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, i.e. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organization of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterized by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.

XX Sequence 1404 AA:

Query Match 100.0%; Score 5416; DB 13; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 2,6e-298;  
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKNNKKRRTKKKPPKPVYDEAGSLNDGDFVTPPTSTTHNNKVSPTKITTAKPIN 60  
DB 200 vknnkkrrtkkkppkpvvdeagsgldngdfvtpptstthnnkvsptkittakpin 259  
QY 61 PRPSLPNSDTSKETSILTNNKETTVEKETTNNKQTSIDGKEKTSAKTOSTIEKTSAK 120  
DB 260 prpslpnstdtsketsiltnnkettvetketttnnkqtsidgkekttsaketsaksak 319  
QY 121 DLAPTSTVLAKEPTTKATTTGKALTPPKKEPTPTTKKEPASTTPPKKEPTTTISAPPTPK 180  
DB 320 dlaptstvlakpttkatttgkaltppkkeptpttkkepasttppkkeptttisapptpk 379  
QY 181 EPAPTTSKAPTPPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240  
DB 380 epapttskaptppkkepapttttkepapttttkepapttttkepapttttkepaptttt 439  
QY 241 PKKPAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 300  
DB 440 pkkpapttpkepapttttkepapttttkepapttttkepapttttkepapttttkepapt 499  
QY 301 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360  
DB 500 ptttkepapttttkepapttttkepapttttkepapttttkepapttttkepapttttke 559  
QY 361 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420  
DB 560 papttpkepapttttkepapttttkepapttttkepapttttkepapttttkepapttt 619  
QY 421 TPKKLPTTPPKLAPTTPEKAPPTTPEELAPTTPEEPPTTPEEPAPPTTPEKAAAPNTTPE 480  
DB 620 tpkklpttpkklapttpekapttpeelapttpeeppttpeepappttpekaaapnttpe 679

QY 481 PAPITPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 540  
DB 680 papitpkkepapttttkepapttttkepapttttkepapttttkepapttttkepaptttt 739  
QY 541 EPTSTTSKAPPTTPKAGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600  
DB 740 eptsttskappttpkagapttttkepapttttkepapttttkepapttttkepaptttt 799  
QY 601 KELAPTTTSGPTSTSDKAPATTPEKETAPTTTPKEPAPTTTTPPKKAPPTPEPTTSEVST 660  
DB 800 kelaptttsgptstsdkapattpeketaptttpkepapttttppkkapptpepttsevst 859  
QY 661 PTTTKEPTTHKSPESPPELSAETTPKALENSKPEPEVPTTTPKPAATKPEMTTAKDKT 720  
DB 860 ptttkeptthkspesppeelsaettpkalenskpepevptttpkpaatkpemttakdk 919  
QY 721 TERDLRTPEPTTTAAPKMTKETATTTETTESKITATTTQVSTSTTTQOTTPPKITTLKT 780  
DB 920 terdlrtpeptttaapkmtketatttetteskitatttqvststttqottppkittlkt 979  
QY 781 TLAPVTTTTKITTTTELINNKREETAKPKDRTNKSATTPKPKPKPTSKPKPT 840  
DB 980 tlavttttkittttelinnkreetakpkdrtnskattppkpkpkptskpkpt 1039  
QY 841 MPVRKPKTTPPRKMTSNMPLNPTSRIAEAMLOTTTRPNQTPNSKLVENPKSEDAAG 900  
DB 1040 mpvrkpkttpprkmtsnmplnptsriaeamlotttrpnqtpnsklvenpkseadag 1099  
QY 901 AEGEPHMLLRPHVEMPEVTPMDYLPRVNOGIIINPMLSEPTINCNKPKVDGLTTLRN 960  
DB 1100 aegephmlrrphvempevtpmdylprvnoგიიინპლსეპტინცნკპვდგლტლრნ 1159  
QY 961 GRTVAFRGHYFMWLSPESPSPARRTEVWGISPIDVYFTRCNCEGKTFEKK 1013  
DB 1160 grtvafrghyfmwlspepsparritevwigispidvyftrcncegktfek 1212

# RESULT 3

AAB60568  
ID AAB60568 standard; Protein: 1404 AA.

AC AAB60568;  
XX  
DT 27-APR-2001 (first entry)  
XX

DE Human megakaryocyte stimulating factor (MSF, CACP).

XX Human; CACP protein; campodactyl-arthropathy-coxa vara-pericarditis;  
KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopath;c;  
KW antiarthritic.

OS Homo sapiens.

PN WC200107068-A1.

XX 01-FEB-2001.

PF 21-JUL-2000; 2000MC-US20002.

XX 23-JUL-1999; 99US-0145328.

PR 19-JUL-2000; 2000US-0145328.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

PI Warman ML;

XX WPI: 2001-182721/18.

PT New composition comprising the campodactyl-arthropathy-coxa  
PT vara-pericarditis protein in combination with an anesthetic, useful for  
PT treating osteoarthritis, or as lubricants of tissue and joints

DR N-PSDB: AAH98981.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 PS Claim 20: Page 1198-1201; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 CC  
 XX Sequence 1299 AA:  
 SQ  
 Query Match 100.0%; Score 5416; DB 22; Length 1299;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-298;  
 Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VKDKKKNNRTKKKPKPPVNDAGSGLDNGCFKVTTPDTSTTQNKVSTSKITTAKPIN 60  
 DB 200 VKDKKKNNRTKKKPKPPVNDAGSGLDNGCFKVTTPDTSTTQNKVSTSKITTAKPIN 259  
 QY 61 PRSLPNSDTSKESLTJVNKEITVETKETTNTKOTSTDGKEKTSKAKTOSIKTSK 120  
 DB 260 PRSLPNSDTSKESLTJVNKEITVETKETTNTKOTSTDGKEKTSKAKTOSIKTSK 319  
 QY 121 DLATSVLVAKKPTPKAKATTGKALTPKKEPTTPPKKPASTTPKEPTTTSAPTPK 180  
 DB 320 DLATSVLVAKKPTPKAKATTGKALTPKKEPTTPPKKPASTTPKEPTTTSAPTPK 379  
 QY 181 EPAPTTKSAATTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 240  
 DB 380 EPAPTTKSAATTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 439  
 QY 241 PKRAPPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTP 300  
 DB 440 PKRAPPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTP 499  
 QY 301 PTPPKKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 360  
 DB 500 PTPPKKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 559  
 QY 361 PAPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 420  
 DB 560 PAPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 619  
 QY 421 TPKKLTPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTP 480  
 DB 620 TPKKLTPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTP 679  
 QY 481 PAPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 540  
 DB 680 PAPTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 739  
 QY 541 EPSTSTSDKRAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 600  
 DB 740 EPSTSTSDKRAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAP 799  
 QY 601 KELAPTTTGGPTSTSDKRAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 660  
 DB 800 KELAPTTTGGPTSTSDKRAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPTTTPKEAPPT 859  
 QY 661 PTTTKKPTTTHKSPDESTELSAEPTPKALENSPKPGVPTTKTPATPEMTTAKDKPT 720  
 DB 860 PTTTKKPTTTHKSPDESTELSAEPTPKALENSPKPGVPTTKTPATPEMTTAKDKPT 919  
 QY 721 TERDLATTPETTTAAKMTKETATTTETKESKITATTTQVSTTTQDTTPPKITTLKTT 780  
 XX

DB 920 terdlrttpebttaapkmketaltektlesklatlattqvtsttqdtctpfkittlktc 979  
 QY 781 TLAPKVTYTKKTTTITTEIMNKPEETAKPKDRAATSKATTPKPOKPTAKPKPTSTKKPK 840  
 DB 980 tlapkvttkktlittetmknpeetakpkdratnskattpkppokptakpkptstkkpk 1039  
 QY 841 MPVRKRTKTPTPKPKMSTYBMLNPSTRIAEAMLQTTTRPNQTPNSKLVEVNPSEADAG 900  
 DB 1040 mpvrkrtktpkpkmtstymblnpstriaeamlqtttrpnqtpnslvevnpksedag 1099  
 QY 901 AEGTTPMLLRPHVEMPEVTPDMDYLRVNPNOGIIINPMLSDEFNINCGRPVGLTTLRN 960  
 DB 1100 aegttppmlrrphvempevtpdmdylrvnpnogiiinpmldsdefnincgkpvgltlrn 1159  
 QY 961 GTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSIDVTFRNCNCKGKTFPK 1013  
 DB 1160 gtlvafrghyfwmllspspsparritevwwgipsidvtfrncncgkktffk 1212  
 RESULT 2  
 AAR26049  
 ID AAR26049 standard; Protein; 1404 AA.  
 XX  
 AC AAR26049;  
 XX  
 DT 02-FEB-1993 (first entry)  
 XX  
 DE MSF precursor.  
 XX  
 KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
 XX stability; proteolytic cleavage; adhesion; alternative splicing.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers.  
 FT Region 1..26  
 FT /label= Exon\_I  
 FT 26..67  
 FT /label= Exon\_II  
 FT 67..107  
 FT /label= Exon\_III  
 FT 107..157  
 FT /label= Exon\_IV  
 FT 157..200  
 FT /label= Exon\_V  
 FT 200..1141  
 FT /label= Exon\_VI  
 FT 1411..1166  
 FT /label= Exon\_VII  
 FT 1166..1212  
 FT /label= Exon\_VIII  
 FT 1213..1266  
 FT /label= Exon\_IX  
 FT 1266..1331  
 FT /label= Exon\_X  
 FT 1331..1373  
 FT /label= Exon\_XI  
 FT 1373..1404  
 FT /label= Exon\_XII  
 FT  
 PN W09213075-A.  
 PD 06-AUG-1992.  
 XX  
 PE 17-JAN-1992; 92WO-US00433.  
 XX  
 PR 18-JAN-1991; 91US-0643502.  
 PR 10-SEP-1991; 91US-0757022.  
 XX  
 PA (GENY ) GENETICS INST INC.  
 XX  
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
 XX

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:33 ; Search time 107.17 Seconds  
(without alignments)  
700.161 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1212

Perfect score: 5416  
Sequence: 1 VKDNKKRKRRKKRPKPPVY.....SPIDIVFRNCGEKGTFPRK 1013

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Genecod\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5416	100.0	1299	22	AAAM24322 Human EST encoded
2	5416	100.0	1404	13	AAAR26049 MSF precursor.. Sy
3	5416	100.0	1404	22	AAAB60568 Human megakaryocyt
4	5416	100.0	1404	22	AAAB29773 Human megakaryocyt
5	3217.5	59.4	902	22	AAAB29778 Human MSF-derived
6	1357.5	25.1	452	16	AAAR80041 Human megakaryocyt
7	1168	21.6	5179	22	AAAM24516 C899P predicted am
8	981	18.1	763	21	AAAG38942 Arabidopsis thalia
9	942	17.4	1664	19	AAWA3106 C. thermocellum Ol
10	823.5	15.2	472	22	AAAB60569 Bovine MSF ortholo
11	771.5	14.2	1325	22	AAAM03645 Peptide #2327 enco

12	715	13.2	763	18	AAW31852
13	625.5	11.5	4412	21	AAV53666
14	612	11.3	572	18	AAW31855
15	544	10.0	844	7	AAAP60570
16	542	10.0	807	21	AAV54467
17	520.5	9.6	788	21	AAV54466
18	508	9.4	1837	21	AAAB11726
19	506.5	9.4	744	9	AAAP82975
20	486.5	9.2	2971	21	AAAB41231
21	489.5	9.0	2972	22	AAAB50363
22	489.5	9.0	3118	22	AAAB50362
23	489.5	9.0	1721	21	AAAB11727
24	488.5	9.0	826	13	AAAR26042
25	488	9.0	617	22	AAAM4187
26	488	9.0	617	22	AAAM4187
27	488	9.0	957	21	AAV59288
28	488	9.0	957	22	AAW4513
29	488	9.0	1721	22	AAW48299
30	485	9.0	1127	22	AAAB95541
31	476.5	8.8	652	9	AAAP82974
32	467.5	8.6	511	22	AAAM14883
33	467.5	8.6	511	22	AAAM27312
34	467.5	8.6	511	22	AAAM2607
35	450.5	8.3	378	12	AAAR14160
36	446.5	8.2	378	12	AAAR14162
37	446.5	8.2	750	20	AAAY05477
38	443	8.2	2870	21	AAAY95559
39	443	8.2	3178	21	AAAY95556
40	442.5	8.2	751	16	AAAR80839
41	433	8.0	2819	22	AAAB35408
42	428.5	7.9	2665	22	AAAM14533
43	428.5	7.9	2665	22	AAAM26950
44	428.5	7.9	2665	22	AAAM02259
45	428.5	7.9	3266	21	AAAB42491

## ALIGNMENTS

RESULT 1	
ID	AAAM24322 standard: Protein; 1299 AA.
XX	AAAM24322;
DE	12-OCT-2001 (first entry)
XX	Human EST encoded protein SEQ ID NO: 1847.
XX	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	WO200154477-A2.
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001MO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
XX	
PR	17-JUL-2000; 2000US-0617746.
XX	
PR	03-AUG-2000; 2000US-0631451.
XX	
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX	
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the Pa  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precu  
Human ORFX ORF995  
Human SRCAP. Homo  
Portion of CRYPTOS  
P. yoelii SSP ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 polype  
C900P predicted am  
Cryptosporidium pa  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
PRP 378. Triflicum  
PRP encoded by clo  
C. albicans Rb1 p  
Caenorhabditis ele  
Caenorhabditis ele  
Japanese sea musse  
Human 07CG27 gene  
Peptide #967 enco  
Peptide #987 enco  
Peptide #941 enco  
Human ORFX ORF2255

Mon Apr 29 08:35:38 2002

us-09-556-246-1\_copy\_200\_1263.rag

2 - 3

Page 16



```

Db 3948 ap---eeelapeekpvpvasee-----eeevpppavpeepkklipkvp-vikkpeapp 3999
QY 595 PKKPAKELATTTKGGPSTSDKPAKTPKETA-----PTPKKEAPPTPKKP--AP 645
  |||:::|
Db 4000 kpepepekv---lekpkikpppppppapedvkekifglkaipkpkvpenpqppeke 4056
QY 646 TTPETPPPTTSEVS--TPTTKKP-----TTIHKSPDESTPPLSAEPTRKALENSPK 696
  |||:::|
Db 4057 ltpklvpgeekvkrlllperkpepkeevlksvlrkrpeeeepkve---pkkie-kvkk 4111
QY 697 PGVPTTKPATKRP-----EMTTAKOKTTEEDLTTPETTAAKMKETATTTKTTES 752
  |||:::|
Db 4112 pavp---epppkpveevepvlvkeriklpeclkvpeikpaplpaepkpkpea---- 4164
QY 753 KITATTTQVSTTQDTPPTTKTTTLAPKVTTKTKTITTEIKNK-----PETA 806
  |||:::|
Db 4165 -----evklkppvpepelpiaapvtpvvgkkaekapkeea 4204
QY 807 KPKRATNSKATTPPKPKPKPKPKPS--TKKPKTPRVRKPKTPTPP-----RKMTST 859
  |||:::|
Db 4205 kpkpplkvppkkipspiaeeriklirpsgeekppdeapfeyqlkavplkfvkeikdlilt 4264
QY 860 MPELPTSRIAEAMLOTTT 878
  |||:::|
Db 4265 esetvgsaalfectvpspt 4283

```

## RESULT 14

AAWJ1855  
ID AAWJ1855 standard; Protein; 572 AA.

XX AAWJ1855;

DT 27-APR-1998 (first entry)

XX Mycobacterium tuberculosis 55 kDa protein.

XX Tuberculosis: mycobacteria: infection; diagnosis;

KM antimycobacterial; antibiotic; vaccine.

XX Mycobacterium tuberculosis.

XX W09741252-A2.

PD 06-NOV-1997.

PF 18-APR-1997; 97WO-EP01973.

PR 29-APR-1996; 96DE-4017184.

PA (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.

PI Espilita C, Honisch C, Moreno C, Singh M;

DR WPI: 1997-549750/50.

DR N-PSDB; AAT93610.

XX New DNA and related proteins or RNA derived from M. tuberculosis  
PT used for diagnosis of mycobacterial infections, monitoring  
PT vaccination and development of anti-mycobacterial agents

PS Claim 11; Fig 16; 55pp; English.

XX This novel 55 kDa protein is encoded by an open reading frame of  
CC a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing  
CC polymorphic GC-rich sequences. Its amino acid sequence shows  
CC a high proline content, but there is no homology to any known  
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis  
CC proteins (see AAWJ1851-57) are claimed. These can be produced as  
CC recombinant proteins, especially in bacterial, yeast, fungal or  
CC higher eukaryote host cells, and used for diagnosing tuberculosis  
CC and other mycobacterial infections in humans or animals. The  
CC claimed proteins can also be used for epidemiological studies, for

CC monitoring vaccination, and for the development of vaccines and  
CC anti-mycobacterial drugs.

XX Sequence 572 AA;

Query Match 10.7%; Score 612; DB 18; Length 572;  
Best Local Similarity 30.6%; Pred. No. 1.8e-27;  
Matches 193; Conservative 29; Mismatches 253; Indels 156; Gaps 33;

```

QY 153 PTTKEPASTPKKEPTTTTTSKAPTTKE-----PATTTKSAPTTKKEAPTTTKEPAP 207
  |||:::|
Db 2 paapkskpsasppap-----pmpacpmetlpllpvppdlisketcpapppaplpavp 57
QY 208 TTPKEPAPTTTKEPAPTTTTSKAPTTKKEPAP-----TTKPPAPTTTKEPAPTTTKE 259
  |||:::|
Db 58 lppvpplp-----pvpnkilppap-----papvavaavlvapoppilplpnhppapaa 107
QY 260 PPTTKEPAPTTTKEPAPTTTKEPAPTTKAPAKKAP-----TT----- 295
  |||:::|
Db 108 pvpqvlaplpnshppap--psapvpqvlaplpisqrvsvwksqfllslfcccrcvsg 165
QY 296 -----PKEPA-----PTTKEPAPTTTKEPSTTKEPAPTTTTSKAPTTTKEPAPTT 342
  |||:::|
Db 166 evlagalnpssrsplltltpalpai-----pplpllpplintavppl---pplpv 217
QY 343 TKSAPTTPK-EPSPITTKKP--APTTPKEPAPTTKPKKAPTTTKEPAPTTTKEPAPTT 399
  |||:::|
Db 218 talaplpilaplpisqrvpappilppqkpwclpplapapepk-tpvllppqscppse 276
QY 400 KPAPTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 457
  |||:::|
Db 277 knpnpap--peppeksspalppapppasmpsaavrvpsplppap-----paapras 327
QY 458 TPTTPEEPAPTTPKA--AAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 511
  |||:::|
Db 328 mpalp--papppapalrclpplpssppapnp--papap---tlpklklsanpccpv 379
QY 512 GTAPTTLKRPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 571
  |||:::|
Db 380 ppapn--lppapap--pappel-----pappdp--pupvansppap 416
QY 572 KEAPPTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 624
  |||:::|
Db 417 --papppapsalp-fvnpappplpapksrpalpaapppappparatlpppapppap 473
QY 625 KETA--PTTKEPAP--TTKPKAPPTTPEPTTSEVSTPTTKEPTTIHKSPDESTP 679
  |||:::|
Db 474 nsmaipppapppilpilaipapppaplpmsppapppilppaapppapppilpnpp---sp 530
QY 680 ELASPTPKALENSPKKEPVPPTTKTPATKRP 710
  |||:::|
Db 531 pla-----pvpqaplaplpingrp 549

```

## RESULT 15

AA60570  
ID AA60570 standard; Protein; 844 AA.

XX AA60570;

DT 24-AUG-1991 (first entry)

XX Sequence of the Falciparum Interspersed Repeat Antigen  
DE (FIRA).

XX Malaria vaccine; antigen; epitope.

OS Plasmodium falciparum.

XX W08601802-A.

PD 27-MAR-1986.

```

OY 306 EP--APTTTKEPSTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPSPSTTKEPAP 363
Db 149 fppfpapalppap-----paplaanspplppapppagtl--ppaawpvyppaapkskspa 201
OY 364 TTPKEPAPTTPKKAPPTTTPKEPAPTTPKEPAP-----TTTKEPAPTTPKEPAPTTPKETAP 419
Db 202 spvpppap-----pmatpmeifpplpvpvpdpisketppapppapppapppapppapppap 256
OY 420 TTPKLTPTTEKLAAP-----TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 469
Db 257 pvpnlipppap--appvavaavlvapcpplpplpnhppapppaavpvpplaplnshp 313
OY 470 ---PKAAPNTPKKEPAP----- 483
Db 314 pappapavpvpplapplsipvsvwksftllstfcrcvsgvlagalnpssrpspl 373
OY 484 ---TTPKEPAPTTPKEPAPTTPKEPAP-----PTTKEPAPTTP--KAPAKE 533
Db 374 tttppalpaplpplpplpplntavpplpplpvtalappplapppapppapppapppap-- 431
OY 534 LAPTTTKEPSTTSKAPAPTTPKGTAPTTKKEPAP-----PTTKEPAPTTPKGTAPTTTL 587
Db 432 --pippgkpwltplapapppk--lvplpvgpseppsekmpnpappppppppksspalpp 488
OY 588 KKPAPTTP--KKAPAKELAPTTTGP--TSTTSKAPAPTTPKET--APTTPKEPAPTTP 640
Db 489 appapmpsaavrvppspplpplpappaprasmpalppappspatrlcpplpsspapnsp 548
OY 641 KKPAPTTPPTPTTSEVSTPTTKEPTTIHKS-----DESPPELSAETPKA--LENSP 694
Db 549 --pappaplpkplis--anppcovpvpapnpppappapppapppapppapppapppap 604
OY 695 KEGVPTTTPATKPKEMTTAKDTERDLRTPEPTTAAPKMKETATTEKTESKI 754
Db 605 --pappapppapppapppapppap-----ppcpaapk-----srpal 637
OY 755 TATTTOVSTTTQDTPPKITTLKTTTLPARKVTTTKTTTTEIMKPKPEAKPKMRATN 814
Db 638 paappapppapppapppap-----ppappapppap 665
OY 815 SKATTPKPKP-----TKAPKKTSTKPKPMPRKKTTPPKMTSTM-----PELNP 865
Db 666 smalpapppppplpplpappapppapppapppapppapppapppapppapppapppap 725
OY 866 TSRIAEAMLOTTRPNQTPNSKLVENPKSEDAEGAEGT 905
Db 726 vpgaplaplpingrvfarknsl-----gsssgdt 756

```

```

PA (OVAR-) QUARK BIOTECH INC.
XX Elmat P, Mor O, Skallier R, Feinstein E, Faerman A:
XX WPI; 2000-053304/04.
XX Identification of stress induced genes for determining risk and
XX preventing, treating or controlling osteoporosis
XX Claim 32; Fig 6A-R; 308pp; English.
XX The present sequence is obtained from a clustal X alignment with
XX protein 608. Protein 608 was identified using the method of the invention
XX after subjecting rat osteoblasts to mechanical stress. Expression of the
XX 608 gene was found to be upregulated by about 3-fold in cells subjected
XX to mechanical strain. The specification describes a method for the
XX identification of genes responsive to a specific mechanical stress. The
XX method comprises applying the mechanical stress to an organism (tissue
XX or cells comprising bone cells), isolating the specific cellular
XX fractions and extracting mRNA from them, and differentially analysing the
XX mRNA in comparison with control samples. The method is used to identify
XX genes whose expression is responsive to a specific stress. The identified
XX genes are employed in determining risk associated with a physiological or
XX disease state. The risk determination methods are used for testing a
XX medicament for gene therapy. These medicaments, or genes identified by
XX the method of the invention, are used for treating, preventing or
XX controlling a physiological or disease state (especially osteoporosis or
XX bone density) or other factors causing or contributing to osteoporosis or
XX its symptoms or other conditions involved in mechanical stress or its
XX lack. The methods can also be used for advancing research or studies in
XX bone development.
SO Sequence 4412 AA:
Query Match 11.0%; Score 625.5; DB 21; Length 4412;
Best Local Similarity 26.2%; Pred. No. 2.5e-27;
Matches 225; Conservative 86; Mismatches 379; Indels 169; Gaps 40;
OY 99 TDGKKEKTSKAKETOSIEKI-----SAKDLAPTSTKVLAKPPKAETTTKGPALN----- 146
Db 3515 tdfgrllgrveehvvekvhyrvfeaevevefckpapp-----kgeisekllp 3566
OY 147 TPKKEPTTP--KEPASTPKKEPTPTTTSAPTTKEP--APTTPKAPPTTPKAPATTTKE 204
Db 3567 pkppkpvkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvp 3621
OY 205 PAPTTPKEPAPTTPKAPATTTTSAPTTKAPATTTPKAPATTTPKAPATTTPKAPATTTPK 260
Db 3622 ---ekkvvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvp 3672
OY 261 TPTTKEPAPTTPKAPATTTP-----KEPAPTTPKAPATTTPKAPATTTPKAPATTTPK 313
Db 3673 kaavpkkvpealppkpepppvefeepsepsapkp--kvpvppvrvpvpkvevpkvp 3731
OY 314 EPSPTTPKEPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATT 373
Db 3732 paap--pkkvpvvpk--vpeapkvvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpk 3782
OY 374 PKPAPTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATTTPKAPATT 429
Db 3783 pekkvp-----ealppkpespppvefeepaealeppaevveeppaapppvvpknpv 3838
OY 430 PEKLAPTT-----PEKAPATTPEELAPTTPEEPPT--TTPKEPAPTTPKAAAPNTPKAPAP 483
Db 3839 pekkapavvakpelpvpvkvpevpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvp 3896
OY 484 ---TTPKEP-----APTTPKEPAPTTPKETAATTTKGTAPTTLKAPATTTPKAPAKE 534
Db 3897 kkvavpkpvpvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvpkvp 3947
OY 535 APTTKEPTSTTSKAPAPTTPKGTAPTTTPKAPATTTPKAPATTTPKGTAPTTTKEPAPT 594

```



```

Db 60 ntrrhskpldktgnskldlh-seltdhheapptseensngkdpmlrngsvdpadst 118
Qy 56 A-----KPINRPSLPPN---SDTSKER-----SLVNNKETVETKER 90
Db 119 tthkesagkhilcpapkskincrstskstgskdctgrpleksmslaktststkt 178
Qy 91 TTT-----NKQSTDGKERTTSAKER-----OSIKET--SAKDLPYTSKVLAKPT 133
Db 179 ttfsmngsqgckqstafpekltlaasktyktgtfpeesektedstvtssdkllkctt 238
Qy 134 PKAETTTGCPALT-----TPKEPTPT---TPKEPASTTPKEPTPT 170
Db 239 knlgecllsanelqslaeptehggrtanennlpsaeptenretlanentlpsaepten 298
Qy 171 TIKSA--PTTPKEBAPT-----TTKSAPTTPKEBAPTTPKE---PATTPKEBAPT 216
Db 299 reretanentlfpagptenretlanentlflpsaeptehggrtanentlpsaeptehggr 358
Qy 217 TTKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 271
Db 359 lane-----ntlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 414
Qy 272 -TKBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 324
Db 415 enertanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 474
Qy 325 -----PTTTPKA-PTTTPKEBAPTTPKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 367
Db 475 qtrpfanektssaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 532
Qy 368 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 422
Db 533 engdrtplaneaktlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 589
Qy 423 KILPTTPPEKLAAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 479
Db 590 speeptenretlanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 649
Qy 480 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 533
Db 650 eptengqtrpfanektssaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 709
Qy 534 LAPTTKEPTSTSDKAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 588
Db 710 aeptenretlanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 769
Qy 589 EPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 621
Db 770 saeptehggrtanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 829
Qy 622 TTPKET-----APTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 670
Db 830 aeptenretlanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 889
Qy 671 HKSPDE-----STPELSAEPPTPK-----ALENSPKPEGVPT-----TKTPA 706
Db 890 saeptehggrtanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 948
Qy 707 ATP-----EMTTAKDKTTERDLRTTPETTTAPKMTKE-----TATTEKTE 751
Db 949 saeptehggrtanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 1008
Qy 752 SKTTAAT--QVNTSTTODTTPFKI-----TTL-----KTTTLAPK 785
Db 1009 spakptehemrpsanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 1068
Qy 786 VTTT-----KTTTTEIMNKPEETAKPKDRATNSKATTPKQPK----- 825
Db 1069 ltpslakptehggrtanentlpsaeptehggrtanentlpsaeptehggrtanentlpsaeptehggr 1128
Qy 826 -----TKAPKPTSTSKKPTTPKPK-----KTTTPPKMTSTMPKL 863
Db 1129 ekmtqvtektstehptktste-ktlrtpckptlysekltckkgnktpvpektlenlgn 1187

```

```

Qy 864 NPTSRIAEMIQTTTPRNOT 883
Db 1188 tlletlkapvkslenpekt 1207

RESULT 12
AAW31852 standard; Protein; 763 AA.
AAW31852:
AAW31852:
27-APR-1998 (first entry)
DE Mycobacterium tuberculosis 74 kDa protein.
KW Tuberculosis; mycobacteria; infection; diagnosis;
KM antimycobacterial; antibiotic; vaccine.
XX Mycobacterium tuberculosis.
OS
XX
XX WO9741252-A2.
XX
XX 06-NOV-1997.
XX
XX 18-APR-1997; 97WO-EP01973.
XX
XX 29-APR-1996; 96DE-4017184.
XX
XX (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX
XX Espilitia C, Honisch C, Moreno C, Singh M;
XX
XX WPI; 1997-549750/50.
XX
XX N-PSDB; AAT93610.
XX
XX New DNA and related proteins or RNA derived from M. tuberculosis -
XX used for diagnosis of mycobacterial infections, monitoring
XX vaccination and development of anti-mycobacterial agents
XX
XX Claim 5; Fig 13; 55pp; English.
XX
XX This novel 74 kDa protein is encoded by an open reading frame of
XX a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing
XX a polymorphic GC-rich sequence. Its amino acid sequence shows
XX a high proline content, but there is no homology to any known
XX proline-rich antigens of mycobacteria. Novel M. tuberculosis
XX proteins (see AAW31851-57) are claimed. These can be produced as
XX recombinant proteins, especially in bacterial, yeast, fungal or
XX higher eukaryote host cells, and used for diagnosing tuberculosis
XX and other mycobacterial infections in humans or animals. The
XX claimed proteins can also be used for epidemiological studies, for
XX monitoring vaccination, and for the development of vaccines and
XX anti-mycobacterial drugs.
XX
XX Sequence 763 AA:

Query Match 12 5%; Score 715; DB 18; Length 763;
Best Local Similarity 28.2%; Pred. No. 3,4e-33;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

Qy 132 PTPKAETTTGCPALTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 191
Db 3 pvp-----aprtalaplpppapppapkskppfppap-----pappcmllvsadp 46
Qy 192 TTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 251
Db 47 pcp--pap-----papkpkskafpfpvpapparelappip--pap-----peapre 90
Qy 252 PAPTTTPKEPTPT-----TPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKE 305
Db 91 srpalpccppppvvlpdppeapppvpapnspfpfpfppeaklvppapvp--pvpnsp 148

```



```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 17.2%; Score 981; DB 21; Length 763;  
 Best Local Similarity 38.5%; Pred. No. 2.6e-48;  
 Matches 226; Conservative 24; Mismatches 293; Indels 44; Gaps 11;

```

QY 137 ETTTGGPALTTPKKEPTTPPKKEPASTTPPKKEPTTTTTSAPTTPKKEPAPTTPKKE 196
DB 126 qteqgnrlylrpprrlpppcvccp-egsp-----pppvlct--pqlpvt 170
QY 197 PATTTPKKEPAPTTPKKEPAPTTPKKEPAPT--TKSAPTTPKKEPAPTTPKKEPAPT 254
DB 171 lrpitlppelpprlpqlpqlpqlpqlpqlpqlpqlpqlpqlpqlpqlpqlpqlpqlp 226
QY 255 TTPKKEPTTPKKEPAPT--KEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 313
DB 227 nuppcupltpclppltpclppltpclppltpclppltpclppltpclppltpclppltp 286
QY 314 EPSPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 373
DB 287 etepnclpelsptclppltpclppltpclppltpclppltpclppltpclppltpclppl 340
QY 374 PKKPAATTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 433
DB 341 ppgtppclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppl 400
QY 434 APTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 493
DB 401 lnpclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltp 460
QY 494 PKKPAATTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 553
DB 461 pklp-pltclppltpclppltpclppltpclppltpclppltpclppltpclppltpcl 514
QY 554 TPKGTAATTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 613
DB 515 tppgclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltp 574

```

```

QY 614 TTSKDP-----APTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKE 666
DB 575 lppclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltp 634
QY 667 P---TTTHKSPDESTEPKLENSPTKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT 710
DB 635 pppclppltpclppltpclppltpclppltpclppltpclppltpclppltpclppltp 681

RESULT 10
AAW43106
AAW43106 standard; Protein; 1664 AA.
XX
AC AAW43106;
XX
XX 16-OCT-1998 (first entry)
DE C. thermocellum O1pB protein.
XX
XX Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;
XX cellulosome integrating protein; scaffoldin dockerin binding protein.
OS Clostridium thermocellum.
XX
FH Key location/Qualifiers
FT Domain 28..192
FT /note="cohesin type II domain"
FT Domain 207..363
FT /note="cohesin type II domain"
FT Domain 409..565
FT /note="cohesin type II domain"
FT Domain 607..763
FT /note="cohesin type II domain"
XX
XX FR2748479-A1.
XX
XX 14-NOV-1997.
XX
XX 10-MAY-1996; 96FR-0005854.
XX
XX 10-MAY-1996; 96FR-0005854.
XX
XX (INSP ) INST PASTEUR.
XX
XX Beguin P, Leibovitz E;
XX WPI; 1998-011569/02.
XX N-PSDB; AAT86623.
XX
XX Cellulase proteins with cohesin or dockerin type II domains - useful
XX for potentiating the activity of multiprotein enzyme complexes
XX
XX Claim 7; Page 31-39; 60pp; French.
XX
XX Multimeric protein, especially enzymatic, complexes are held together
XX by protein-protein interactions between domains designated dockerins
XX and cohesins, which are found on the catalytic and scaffold subunits
XX respectively. An example of such a complex is the cellulose degrading
XX protein complex from Clostridium thermocellum, known as the cellulosome.
XX This complex comprises around 15 proteins including endoglucanases,
XX cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which
XX interact with a central "scaffold" protein designated the cellulosome
XX integrating protein (CIPa; see AAW43108). The catalytic subunits
XX interact with the CIP subunit via conserved 23 amino acid dockerin
XX domains. CIP has been shown to contain 9 copies of a cohesin domain.
XX The invention relates to the isolation of proteins binding to a novel
XX dockerin type domain found in the C-terminal portion of CIP. The new
XX domain is designated a type II dockerin domain (as compared to the type
XX I domain found on the catalytic subunits of the cellulosome). The type
XX II dockerin domain has some sequence similarity to the type I dockerins
XX but is unable to bind type I cohesin domains.
XX The sequence presented here is an example of a protein which binds

```

PR 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140333.  
PR 23-JUN-1999; 99US-0140334.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151081.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 11-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.

```

Db      1770 fspsttttptpcvplcnwgwidsgnfnfhkpgdgdtellgdcvgpwaaaniscratmvp 1829
Qy      428 -----TPEKLAPTT 437
Db      1830 dvpigdqgtvcvdsvglcknedqkpgyvlpmafclnyelnwqceocvqlmtttc 1889
Qy      438 PEKPAPTPBELAPTPEEPPTPT--TPEEPAATTPKAAPAINTPKRPAATTYTKREP-----AP 491
Db      1890 temprrtcrtcttttvtcprcpbgcgctctcttctttvtcprcptgcqtcttctt 1949
Qy      492 TTPKEPAPT-TPEETADPTTGTAFTTLKEBAPTPPKKPAKELAPTTTKEPTSTSDKP 550
Db      1950 ttcvtcprcptgcqtcttcttcttvtcprcptgcqtcttcttcttvtcprcptgc 2009
Qy      551 AP-TTPKGTAFTTPKKEBAPTPPKKEBAPTTTKEGTAFTTLKEBAPTPPKKPAKELAPTTTK 609
Db      2010 tprtcttcttcttvtcprcptgcqtctt-ctpttcttvtcprcptgcqtcttctt 2068
Qy      610 GPTSTTSDEKAP-TTPKEGAFTTPKKEBAPTPPKKAPATTPEETPEPTSEVSTPTTKEPT 668
Db      2069 tprcptgcqtcttcttcttcttvtcprcptgcqtctt-ctpttcttvtcprcptgcqt 2127
Qy      669 TIHKSPDESIPELASBPTRPALNSKEP-----GVPT-TKTPAATKPEMTTYAK 717
Db      2128 t----tcttcttvtcprcptgcqtcttcttcttcttvtcprcptgcqtcttctt 2184
Qy      718 DKTEEDLR---TTP--ETTAAPKMT---KETATTEKTESKINATTTQVISTTQDT 769
Db      2185 prcptgcqtcttcttcttcttvtcprcptgcqtcttcttcttcttvtcprcpt- 2243
Qy      770 TPRKITTLKTTTLAPRVYTT-KKITTTEIMNKPBEETAKPKDRATNSKATTPPKOKPTKA 828
Db      2244 tp---tctt-ttvtcprcptgcqtcttcttcttcttvtcprcptgcqtcttctt 2299
Qy      829 P-KKPSITKKPKMPVRNRKPKTPTPRKMTSTMPELNP---TSRIEAMIQTTTR-PNOT 883
Db      2300 prcprcptgcqtcttcttcttcttvtcprcptgcqtcttcttcttcttvtcprcpt 2359
Qy      884 PNSKLVEVNPKSEDAGAGETPHMLLRPFVFMPEVT 921
Db      2360 plttcttvtcprcptg---tgpr-cttpttcttvtcpr 2392

RESULT   8
AAB60569
ID AAB60569 standard; Protein; 472 AA.
XX
AC AAB60569;
XX
XX DT 27-APR-2001 (first entry)
XX
DE Bovine MSF orthologue, superficial zone protein (SZP).
XX
KM Bovine; CACP protein; campodactylly-arthropalhy-coxa vara-pericarditis;
KM superficial zone protein; SZP; MSF orthologue; synovial lubricant;
XX osteoarthritis; joint lubrication; osteopathic; antiarthritic.
OS Bos taurus.
XX
PN WO200107068-A1.
PD 01-FEB-2001.
PP 21-JUL-2000; 2000WO-US20002.
PR 23-JUL-1999; 99US-0145328.
PR 19-JUL-2000; 2000US-0145328.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
PI Watman ML;
XX
WR WPI; 2001-182721/18.

```

XX New composition comprising the campitodactyl-arthropathy-coxa  
PT vara-pericarditis protein in combination with an anesthetic, useful for  
PT treating osteoarthritis, or as lubricants of tissue and joints -  
XX  
XX  
PS Example 1; Fig 4; 34pp; English.  
PS  
XX  
XX The invention relates to a method of treating osteoarthritis via the  
CC administration of a composition comprising the campitodactyl-arthropathy-  
CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
CC The composition may further comprise a local anesthetic. The composition  
CC of the invention may be administered via intra-articular or intravenous  
CC injection. The human CACP protein is identified in the invention as  
CC being megakaryocyte stimulating factor (MSF). The gene encoding  
CC CACP protein (MSF) is located on chromosome 1q23-31, and mutations in  
CC this gene are responsible for the heritable disorder campitodactyl-  
CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
CC acts as a synovium lubricant, and can be used to lubricate tissue and  
CC joints in the treatment of osteoarthritis. The composition may be  
CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
CC loss of range of movement or joint damage). The present sequence  
CC represents the bovine orthologue of human MSF, superficial zone  
CC protein (SZP).

Sequence 472 AA;

Query Match	18.3%	Score 1042.5;	DB 22;	Length 472;
Best Local Similarity	76.4%;	Pred. No. 5e-52;		
Matches 197;	Conservative 20;	Mismatches 38;	Indels 3;	Gaps 3;

QY	808	PKDASITSKATTPROCKETAPKAPKPPSTYTKKPKMPRVKPKKTPPPRK-MISTMELNPT	866
Db	76	pxgratnsqvtcpkqkprtkarkpustckprct-prvkprkcttprkcttsamperct	134
QY	867	SRIAEMLOTTTRBNQJPNBSKLEVNPMSEDAAGEGEPHMLRPHVMYRVTQDMOYL	926
Db	135	s-lpeamigtcttprprpnselidvnsenedgaageekrhmifrrpvlprvrgtel	193
QY	927	PRVNOGIITINPMLSDBTNCNCKRPVDGLTTLNGLTVAFRGHYEMWISPSBPSPARRI	986
Db	194	vtvrsqgfglmpmsdctclcmgrpvdgltlrlngllvaftrhytwmltprfprpprrl	253
QY	987	TEVMGIISSIPIDVTFTRNCEGKTFEFGDSQYMFRTNDIADACVPRPIFPGFCGLGOIVA	1046
Db	254	tevvgipsrldvtrftrncnecgkctffkqsgywfndicldagyrkliskgfygllngklva	313
QY	1047	ALSTAKKKNMPSVYFFK	1064
Db	314	alslsaqyksrpsvuyffk	331

RESULT	9
AAG38942	
ID	AAG38942 standard; Protein; 763 AA.
XX	
AC	AAG38942;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 48115.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
KW	
XX	
OS	Arabidopsis thaliana.
XX	
PN	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	



CC Lubricating moiety and has a polypeptide sequence comprising 1-76  
 CC repeats of a motif having at least 50% identity to the sequence KEPAPTT  
 CC (AAB29774). The invention also relates to a nucleic acid encoding a  
 CC human MSF-derived tribonectin; a biocompatible composition comprising a  
 CC human tribonectin for inhibiting tissue adhesion formation; and a method  
 CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
 CC measuring the amount of MSF or its fragment in a biological sample of a  
 CC mammal, wherein an increased amount of MSF compared to a control  
 CC indicates the presence of or predisposition to osteoarthritis.  
 CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
 CC treatment of osteoarthritis, where they may be used for lubricating  
 CC mammalian joints, such as articulating joints of humans, dogs or horses.  
 CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
 CC useful for inhibiting adhesion between two surfaces such as the injured  
 CC tissues of a mammal, where the injury is caused by a surgical insertion  
 CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
 CC tribonectin may be used in gene therapy. The present sequence represents  
 CC a substantial portion of a human MSF-derived tribonectin.

Sequence 902 AA:

Query Match 61.1%; Score 3484; DB 22; Length 902;  
 Best Local Similarity 74.6%; Pred. No. 1.4e-190;  
 Matches 794; Conservative 28; Mismatches 80; Indels 162; Gaps 66;

QY 1 VKDNKKNRTRKKRPVVDAGSGLDNGDFKVTTPDSTTQHNNVSPKITTAKPIN 60  
 DB 1 vkdnknrtkrrkrrprrvvdagsgldngdfkvttpdsttqhnnvspkittakpin 60  
 QY 61 PPSLPSPNSDTSKETSLVYNNKETTVEKTTTNNKQTSDEKKTSAKETOSIEKTSK 120  
 DB 61 ppslpspnsdtsketsklslynnkettvtetktttnnkqtsdsgkelttsaketsiekttsak 120  
 QY 121 DIAPTSKVLAKPTPAETTKGPAITTPKEPPTTPKEPASTTPKEPPTTPKSAPTPK 180  
 DB 121 diaptskvlakptpaettkgpaittpkeppttpkepasttpkepapttkkaptpk 180  
 QY 181 EPAPTTTTSAPTTPKEPAPTTPKEPAPTTPKBPAPTTPKEPAPTTPKSAPTPK 240  
 DB 181 epapttttsaptpkpeaptpkpeaptpkbpaptpkpeaptpkksaptpk 240  
 QY 173 EPAPITKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 228  
 DB 173 epapitksaptpkpeaptpkpeaptpkpeaptpkpeaptpkpeaptpk 228  
 QY 241 PKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 300  
 DB 241 pkkpaPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 300  
 QY 229 -keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 281  
 DB 229 -keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 281  
 QY 301 PTTPEKAPATTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSPPTTPKE 360  
 DB 301 pttpekapattpkepapttpkepapttpksaptpkpeaptpksppttpke 360  
 QY 282 PCT-keap-ctkeap-ctkeap-ctkeap-ctkeap-ctkeap-ctkeap-ctke 335  
 DB 282 pct-keap-ctkeap-ctkeap-ctkeap-ctkeap-ctkeap-ctkeap-ctke 335  
 QY 361 PAPTPPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 420  
 DB 361 paptppkapapttpkapapttpkapapttpkapapttpkapapttpkapapttp 420  
 QY 336 PAPIT-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 388  
 DB 336 papit-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 388  
 QY 421 TPKKLPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLA 480  
 DB 421 tpkkLPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLAPTTPPEKLA 480  
 QY 389 t-keaplttkep-aplttke-papltt-keapltt-keapltt-keapltt-ke 440  
 DB 389 t-keaplttkep-aplttke-papltt-keapltt-keapltt-keapltt-ke 440  
 QY 481 PAPTPPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 540  
 DB 481 paptppkapapttpkapapttpkapapttpkapapttpkapapttpkapapttp 540  
 QY 441 PAPIT-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 488  
 DB 441 papit-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 488  
 QY 541 PPTSTSTSKAPATTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTP 600  
 DB 541 pptststskapattpkapapttpkapapttpkapapttpkapapttpkapapttp 600  
 QY 489 EPAPITKE-papltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 541  
 DB 489 epapitke-papltt-keapltt-keapltt-keapltt-keapltt-keapltt-keapltt-kepa 541  
 QY 601 KELAPTTTTGAPTSTSDAPATTPKETAAPTTPKEPAPTTPKAPAPTTPKAPAPTTP 660  
 DB 601 kelapTTTTGAPTSTSDAPATTPKETAAPTTPKEPAPTTPKAPAPTTPKAPAPTTP 660  
 QY 542 -----tkkeaplttke-papltt-keapltt-keapltt-keapltt-keaplttkepa- 589  
 DB 542 -----tkkeaplttke-papltt-keapltt-keapltt-keapltt-keaplttkepa- 589  
 QY 661 PTTTKEPTTIHKSDESTPELSAETTPKALENSPKREPVPPTTKTPAATKPEMTTAKDKT 720  
 DB 661 ptttkepttiHKSDESTPELSAETTPKALENSPKREPVPPTTKTPAATKPEMTTAKDKT 720

DB 590 -pttkeaplttkeapltt----kepap-----tkkep-aplttkeaplttkeaplttke--- 636  
 QY 721 TERDLRTTETTTAAPKMKKEATTTTEKTTESKITATTQVSTYTTQDTTPRKITTLKTT 780  
 DB 637 -----paptt-----tkkeapltt-----kepaplttkeaplttkepap----- 667  
 QY 781 TIAPKVTTTKITTTTTEIMNKPEETAKPKORATNSKATPKPOKPKAPKKTSTKKPKT 840  
 DB 668 -----tkkeaplttkepap-----tkkeaplttkepap-----tkkeaplttkepap----- 686  
 QY 841 MPKVRKKTTPPPKMTSTMPKELNPTSRIAEAMLQTTTPRNOTPNSKLVENPKRSDAG 900  
 DB 687 -----aplttpprkmtstempelnptsrtaeaml-cttrpnqtlpskltvevnpkrsdagg 739  
 QY 901 AGSEPHMLLRHVPMPETTPMDVLPRVNPNGIITINPMLSDETNCKNGKPDGLTTLRN 960  
 DB 740 aegsephmlLRHVPMPETTPMDVLPRVNPNGIITINPMLSDETNCKNGKPDGLTTLRN 960  
 QY 961 GTLVAFRGHYFWMLSPSPSPARRITEVWGIPIPIDTFTVTCNCEGKTFEKKDSQYMR 1020  
 DB 800 gtlvafrghyfwmLSPSPSPARRITEVWGIPIPIDTFTVTCNCEGKTFEKKDSQYMR 1020  
 QY 1021 TNDIDAGTTPKPIFKFGGGLTQIYAALSTAKYKMWPSVYFEK 1064  
 DB 860 tndidagtpkpiFKfggltqivaalsta-ykmwpsvyeffk 902

RESULT 6  
 AAR80041  
 ID AAR80041 standard; Protein: 452 AA.  
 AC AAR80041;  
 XX  
 DT 10-APR-1996 (first entry)  
 XX  
 DE Human megakaryocytopoietin protein.  
 XX  
 KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;  
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;  
 KW multipotential stem cell.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 393..396  
 FT /note= "unspecified amino acids"  
 FT Misc-difference 444..446  
 FT /note= "unspecified amino acids"  
 XX  
 FN W09523861-A1.  
 PN  
 PD 08-SEP-1995.  
 XX  
 PE 06-MAR-1995; 95WO-CN00015.  
 XX  
 PR 04-MAR-1994; 94CN-0112066.  
 XX  
 PA (SHAN-) SHANGHAI BEITI BIOTECHNOLOGY CO LTD.  
 XX  
 PI Gu X, Han Z, Shen Q;  
 XX  
 DR WPI: 1995-320576/41.  
 DR N-PSDB; AAT04546.  
 XX  
 PT New haematopoietic cell growth factor - used for treating  
 PT thrombocytopenia and hematocytopenia  
 XX  
 PS Example: Page 23; 36pp; Chinese.  
 CC This sequence represents the human megakaryocytopoietin (MPO) protein.  
 CC This sequence was purified using a carrier which can couple wheat germ  
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
 CC AAR80039 and AAR80040) were used to produce the amplification primers

XX Claim 3; Page 7; 47pp; English.

PS The invention relates to a human tribonection which is a product of  
XX alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonection has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76  
CC repeats of a motif having at least 50% identity to the sequence KKPAPT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a  
CC human MSF-derived tribonection; a biocompatible composition comprising a  
CC human tribonection for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonection and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonection, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC tissues of a mammal, where the injury is caused by a surgical insertion  
CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonection may be used in gene therapy. The present sequence represents  
CC human MSF.  
XX  
XX  
SQ Sequence 1404 AA:

Query Match 100.0%; Score 5698; DB 22; Length 1404;  
Best local similarity 100.0%; Pred. No. 0;  
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKNNKNNFKKKKKPPKPPVNDAGSGLDNGDKVTPPTSTTQNNKVTSEKITAKPIN 60  
DB 200 VKdhknktkkkppkpyvdaagsglndgdfkvtlpstltgknhkvtstpkltakpin 259  
QY 61 PRPSLPKSDSKESLTVNKKETTKETTTNKTSDGKEKTSKKEQSIKTSK 120  
DB 260 PRPSLPKSDSKESLTVNKKETTKETTTNKTSDGKEKTSKKEQSIKTSK 319  
QY 121 DLAPTSLVAKPTPKAETTTGPAITTPKEPTTPPKESPASTTPKPAITTKSA 180  
DB 320 dlapskslvakptpkactttgpaalttpepttppkespasttpkpaalttk 379  
QY 181 EPAPPTTSAPPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 240  
DB 380 epapltksaplttkpaplttkpaplttkpaplttkpaplttkpaplttk 439  
QY 241 PKKPAPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 300  
DB 440 PKKPAPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 499  
QY 301 PTPKPAPTTKKESPTTPKPAPTTKKSAPTTKKPAPTTKKSAPTTKKESPT 360  
DB 500 ptkpkeaplttkespttkpkeaplttkkapttkkpaalttkkapttkespt 559  
QY 361 PAPPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 420  
DB 560 paptpkpeaplttkkpaalttkkpaalttkkpaalttkkpaalttkk 619  
QY 421 TPKKLTPTPEKLAPTPEKPAPTPEELAPPTPEEPPTPEEPAPPTPKAA 480  
DB 620 tpkkltpptpeklaptpekpaaptpeelapptpeepptpeepapptpkaa 679  
QY 481 PAPPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 540  
DB 680 paptpkpeaplttkkpaalttkkpaalttkkpaalttkkpaalttkk 739  
QY 541 EPTSTSDKPAPTTKGAPPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 600  
DB 740 eptstsdkaplttkgapttkkpaalttkkpaalttkkpaalttkkpaalttkk 799

QY 601 KELAPTTKGPSTTSKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTK 660  
DB 800 kelaplttkgptsttskpaapttkkpaalttkkpaalttkkpaalttkk 859  
QY 661 PTTTKPEPTTIHKSDESTPELSAEPPTKPALENSKPEGVPTTKPAATKPM 720  
DB 860 ptttkpepttihsdestpelsaepptkpaalenskppegvpttkpaatkpm 919  
QY 721 TERDLRTPEPTTAAPKPKTKETATTTETSKITATTTQVSTTODTPEFK 780  
DB 920 terdlrtpepttaapkpktketatttetskitatvtqvtsttdqtpfk 979  
QY 781 TLAPKVTTKKTTTITTEIMNKEEPAKPKDRATNSKATTPKPKPKSTKKPK 840  
DB 980 tlapkvttkktttiteimnkeetapakpdratnskattpkpkkpkstkkp 1039  
QY 841 MPVRKPKPTTPPKKMTSTMBELNPTSLIAFAMLOTTTPKOTPNKSLIV 900  
DB 1040 mpvrkpkpttppkkmstmbelnptsliafamlotttpkotpnsksliv 1099  
QY 901 AEGEPHMLLRPHVEMPEVTPDMVLPVPRNQGIIINPMLSDETNIGKRP 960  
DB 1100 aegephlmlrphvempevtpdmvlpvprnqgiiinpmlsdetnigkrp 1159  
QY 961 GTLVAFRGHYFWMLSPEFSPSPARITPEVWGISPDIDVFTTRCNCEG 1020  
DB 1160 gtlvafrghyfwmlespefspsparitpevwgispididvfttrcnceg 1219  
QY 1021 TNDIKDAGYPKPIKRGFGGLGQIYALSTAKYKWPSSVYEFK 1064  
DB 1220 tndikdagypkpkirgfgglgqiyaalstakykwpssvyefk 1263

## RESULT 5

AAB29778  
ID AAB29778 standard; Protein: 902 AA.

XX AAB29778;

XX 28-FEB-2001 (first entry)

XX Human MSF-derived tribonection.

XX Human tribonection; MSF; megakaryocyte stimulating factor;

XX alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

XX osteoarthritis; tribosupplementation; tissue adhesion inhibition;

XX friction coefficient reduction; gene therapy; antiarthritic;

XX osteopathic.

XX Homo sapiens.

XX WO200064930-A2.

XX 02-NOV-2000.

XX 24-APR-2000; 2000WO-US10953.

XX 23-APR-1999; 99US-0298970.

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX JAY GD;

XX WPI: 2001-024673/03.

XX Novel tribonection polypeptide useful as lubricant for treating

XX osteoarthritis, comprises O-linked lubricating moiety

XX Disclosure; Fig 1; 47pp; English.

XX The invention relates to a human tribonection which is a product of

XX alternative splicing of the human MSF (megakaryocyte stimulating factor)

XX gene. The tribonection has at least one O-linked oligosaccharide



XX Warman ML;  
 PI  
 DR WPI: 2001-182721/18.  
 XX  
 PT New composition comprising the campodactylly-arthropathy-coxa  
 PT vira-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints -  
 XX  
 PS Example 1: Page -: 34pp: English.

CC The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactylly-arthropathy-  
 CC coxa vira-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactylly-  
 CC arthropathy-coxa vira-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U70316).  
 CC  
 XX  
 SQ Sequence 1404 AA;

Query Match 100.0%; Score 5698; DB 22; Length 1404;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKKRRKKRRKKRPVVDGAGSGDNDGKVTTPDSTTGHKKVSTPKTTAKPIN 60  
 DB 200 VKDKKRRKKRRKKRPVVDGAGSGDNDGKVTTPDSTTGHKKVSTPKTTAKPIN 60  
 QY 61 PRPSLPNSDTSKETSITLVNKETTVETKETTNNKOSTGKREKTSAKTOSIEKTSAK 120  
 DB 260 PRPSLPNSDTSKETSITLVNKETTVETKETTNNKOSTGKREKTSAKTOSIEKTSAK 120  
 QY 121 DLAPTGVLAKKPPKAETTTGGPALTTPKEPTTPPKKPASTTPKKEPTTTIKASAPTPK 180  
 DB 320 DLAPTGVLAKKPPKAETTTGGPALTTPKEPTTPPKKPASTTPKKEPTTTIKASAPTPK 180  
 QY 181 EPATTTKSAPTPKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240  
 DB 380 EPATTTKSAPTPKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240  
 QY 241 PKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 300  
 DB 440 PKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 300  
 QY 301 PTTPKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360  
 DB 500 PTTPKKPAETTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360  
 QY 361 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 420  
 DB 560 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 420  
 QY 421 TPKKLTPPTPKLAPTTTPKPAETTTPEELAPTTPEEPPTTPTEEPAPTTTKEPAPTTTKEPAP 480  
 DB 620 TPKKLTPPTPKLAPTTTPKPAETTTPEELAPTTPEEPPTTPTEEPAPTTTKEPAPTTTKEPAP 480  
 QY 481 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540  
 DB 680 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540

QY 541 EPSTTSKRPAPTTPKGTAETTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 600  
 DB 740 EPSTTSKRPAPTTPKGTAETTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 600  
 QY 601 KELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660  
 DB 800 KELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 660  
 QY 661 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 720  
 DB 860 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 720  
 QY 721 TENDLTPPTTTAABPKKETAATTEKTESKITATTTQVSTTQDTPPKITTLKTT 780  
 DB 920 TENDLTPPTTTAABPKKETAATTEKTESKITATTTQVSTTQDTPPKITTLKTT 780  
 QY 781 TLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKKAPKPKSTKKPKT 840  
 DB 980 TLAPKVTTKKTTTTEIMNKPEETAKPKDRATNSKATTPKPKKAPKPKSTKKPKT 840  
 QY 841 MPVRRPKTTPPRKMTSTMPBLNPTSRIAEAMLTQTPRNOTPNSKLYEVPKSDAGG 900  
 DB 1040 MPVRRPKTTPPRKMTSTMPBLNPTSRIAEAMLTQTPRNOTPNSKLYEVPKSDAGG 900  
 QY 901 AEGETPHMLLRPHVEMPEVTTPMDYLPVPPNOCIIINPMLSDETNINCGRPVDGLTTLRN 960  
 DB 1100 AEGETPHMLLRPHVEMPEVTTPMDYLPVPPNOCIIINPMLSDETNINCGRPVDGLTTLRN 960  
 QY 961 GTLVAFRGHYFWMLSPFSPSPARRTEVWGISPSIDTFTGNCNCGTFFPKDSQYRF 1020  
 DB 1160 GTLVAFRGHYFWMLSPFSPSPARRTEVWGISPSIDTFTGNCNCGTFFPKDSQYRF 1020  
 QY 1021 TNDIKRAGYPRKIFKFGGLTGQIYAALSTAKYKMPESVYEFK 1064  
 DB 1220 TNDIKRAGYPRKIFKFGGLTGQIYAALSTAKYKMPESVYEFK 1064

RESULT 4  
 ID AAB29773 standard; Protein: 1404 AA.  
 XX AAB29773;  
 AC  
 XX  
 XX  
 DT 28-FEB-2001 (first entry)  
 XX  
 DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.  
 XX  
 XX Human MSF, megakaryocyte stimulating factor; tribonectin;  
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
 KW friction coefficient reduction; gene therapy; antiarthritic;  
 KW osteopethic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200064930-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PD 24-APR-2000; 2000WO-US10953.  
 XX  
 FR 23-APR-1999; 99US-0298970.  
 XX  
 PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
 XX  
 PI Jay GD;  
 XX  
 DR WPI: 2001-024673/03.  
 DR N-PSDB: AAC81498.  
 XX  
 PT Novel tribonectin polypeptide useful as lubricant for treating  
 PT osteoarthritis, comprises O-linked lubricating moiety -

PA (GEMV ) GENETICS INST INC.

XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

XX WPI: 1992-284660/34.

DR N-PSDB: MA027223.

XX New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
PT bacterial and viral infections, etc.

XX Claim 1, 2 and 3; Fig 1: 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine preceding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNAs each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V and VI are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, i.e. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organisation of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterised by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.

XX Sequence 1404 AA;

Query Match 100.0%; Score 5698; DB 13; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0;  
Matches 1064; Conservative 0; Indels 0; Gaps 0;

QY 1 VKDNKKRRTKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTOHNVSTSPKTTAKPLN 60  
DB 200 vkdnknrtkkkppkppvvdagsgldngdfkvttpdsttqhkvstspkittakp 259  
QY 61 PRPSLPNSDTSKESLTIVNKKETVETKETTNNKOTSTDGKRTSAAKTOSIEKTSK 120  
DB 260 prpslpnsdtsktseltivnkketvettntknkgtstgktsaktsaktsaktsak 319  
QY 121 DIAPTSVVLAKPKPKATTTKGPALTTPKPEPTTPPKKPASTTPKPEPTTISAPPTPK 180  
DB 320 diaptsvvlakppkpkatttkkgpalttpkpepttpkkpasktppkpeptttisapptpk 379  
QY 181 EPAATTTKSAPTTPKKPAPTTTKREPAPTTKKREPAPTTTKSAPTTKREPAPTT 240  
DB 380 epaapttksapttppkkpaptttkrepapttkkrepaptttkrepaptttkrepap 439  
QY 241 PKKPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTK 300  
DB 440 pkkpapttkkrepapttkkrepapttkkrepapttkkrepaptttkrepaptttkrepa 499  
QY 301 PTTKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTK 360  
DB 500 pttkrepapttkkrepapttkkrepapttkkrepaptttkrepaptttkrepaptttk 559  
QY 361 PAPTTPKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPA 420  
DB 560 papttpkrepapttkkrepapttkkrepaptttkrepaptttkrepaptttkrepap 619

QY 421 TPKKLPTTPEKLAPTTPKREPAPTTPEELAPTTPEEPPTTPEEPAPTTPKKAAPNTPK 480  
DB 620 tpkklpttpeklapttpkrepapttpeelapttpeeppttpeepapttpkkaapntpk 679  
QY 481 PAPTTPKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPA 540  
DB 680 papttpkrepapttkkrepapttkkrepaptttkrepaptttkrepaptttkrepap 739  
QY 541 EPTSTSDKAPAPTTPKKTAAPTTPKREPAPTTKKREPAPTTKKREPAPTTKKREPA 600  
DB 740 eptstsdkapapttppkktaapttpkrepapttkkrepaptttkrepaptttkrepap 799  
QY 601 KELAPTTKGPSTTSKRPAPTTTKKREPAPTTKKREPAPTTKKREPAPTTKKREPA 660  
DB 800 kelapttkgsptsttskrpaptttkkrepaptttkkrepaptttkkrepaptttkrepap 859  
QY 661 PTTKREPTTIKSPDESPELSAEPTRKALENSPKRPGVPTTKTPAATKPEMTTAKDKT 720  
DB 860 pttkreppttikspdespelsaeptrkalenspkrgvpttktpaatkpemttakdk 919  
QY 721 TERDLRTTPETTTAAPKMTKSTATTETKTESKITATTQVYSTTODTTPKRTITLKT 780  
DB 920 terdlrttpeTTTAApkmtkstattetkteskitattqvysttoddtpkrtitlkt 979  
QY 781 TLAPKVTTKTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPKSTKKPK 840  
DB 980 tlapkvttktktitttteimnkpeetakpkdratnskattpkpkptkapkpkstkkpk 1039  
QY 841 MPVRVKKPTTTPPKKMTSTMBELNFTSRIAPAMQTTRPNQOTNSKLVEVNPKSDEAG 900  
DB 1040 mpvrvkpttppkkmtstmbelnftsriapamqtrpnpqotnsklvevnpksdeag 1099  
QY 901 AEGTEPHMLLRPHVMPBEVTPDMOYLPRVPMOGIITNPMI,SDETNICNGKRPVDELTLRLN 960  
DB 1100 aegtephmlrrphvmpbevtpdmoylprvpmogiiitnpmi,sdetnicngkryvdeltlrln 1159  
QY 961 GTLVAFRGHYEWMLSFSPSPARRITEVWGI,PSPIDVFTFRNCCEGTFPKFSQYKRF 1020  
DB 1160 gtlvafrghyefwmlsfspsparritevwgi,pspidvftfrnccegTFPKfsqykrf 1219  
QY 1021 TNDIKDAGYPRPIFKFGGSLTGQIYVALSTAKYKRWMPESYVFEK 1064  
DB 1220 tndikdagypripfkfggsltgqiYvalstakYkRWmpesyvfeK 1263

RESULT 3  
AAB60568 standard; Protein: 1404 AA.  
ID AAB60568;  
AC AAB60568;  
XX 27-APR-2001 (first entry)  
XX Human megakaryocyte stimulating factor (MSF, GACP).  
XX Human: GACP protein; camploactyl-arthropathy-coxa vara-pericarditis;  
XX MSF; megakaryocyte stimulating factor; synovial lubricant;  
XX chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
XX antiarthritic.  
XX Homo sapiens.  
XX OS  
XX WO200107068-A1.  
XX PD 01-FEB-2001.  
XX 21-JUL-2000; 2000MO-US20002.  
XX 23-JUL-1999; 99US-0145328.  
XX 19-JUL-2000; 2000US-0145328.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.

DR N-PSDB: AAH98981.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 PS Claim 20: Page 1198-1201; 1275pp: English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 SS Sequence 1299 AA:

Query Match 100.0%; Score 5698; DB 22; Length 1299;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDKKKKKKKPPKPPVVDGAGSLDNGDFKVTTPDSTGQHNKYSTSPKITTAKPIN 60  
 DB 200 VKDKKKKKKKKKPPVVDGAGSLDNGDFKVTTPDSTGQHNKYSTSPKITTAKPIN 259  
 QY 61 PRESLPNSDTSKETSJLVNKKETTVEKTTTNKQSTDGKEKTSAKTQSIETSAK 120  
 DB 260 PRESLPNSDTSKETSJLVNKKETTVEKTTTNKQSTDGKEKTSAKTQSIETSAK 319  
 QY 121 DLAPTSTKVLAKPPPKAETTKGPAITTPKEPTTPPKKPASTTPKEPTTISAPTPPK 180  
 DB 320 DLAPTSTKVLAKPPPKAETTKGPAITTPKEPTTPPKKPASTTPKEPTTISAPTPPK 379  
 QY 181 EPAPTTTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTP 240  
 DB 380 EPAPTTTKSAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTP 439  
 QY 241 PKKPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKPAAPKPAAPTPKEB 300  
 DB 440 PKKPAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKPAAPKPAAPTPKEB 499  
 QY 301 PTPPKBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTP 360  
 DB 500 PTPPKBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTP 559  
 QY 361 PAPTTPKEBAPTTPKPAAPTPKEBAPTTPKPAAPTPKPAAPTPKPAAPTPKETA 420  
 DB 560 PAPTTPKEBAPTTPKPAAPTPKEBAPTTPKPAAPTPKPAAPTPKETA 619  
 QY 421 TPKKLTPTPPEKLAPTTPKPAAPTPPEELAPTPPEEPPTTPPEAPTPPKAAAPNTPK 480  
 DB 620 TPKKLTPTPPEKLAPTTPKPAAPTPPEELAPTPPEEPPTTPPEAPTPPKAAAPNTPK 679  
 QY 481 PAPTTPKEBAPTTPKPAAPTPKETAAPTPKETAAPTPKETAAPTPKPAAPKELAPTT 540  
 DB 680 PAPTTPKEBAPTTPKPAAPTPKETAAPTPKETAAPTPKETAAPTPKPAAPKELAPTT 739  
 QY 541 EPTSTSDKAPATTPKGAATTPKGAATTPKGAATTPKGAATTPKGAATTPKGAATTP 600  
 DB 740 EPTSTSDKAPATTPKGAATTPKGAATTPKGAATTPKGAATTPKGAATTPKGAATTP 699  
 QY 601 KELAPTTKGPSTTSKAPATTPKETAAPTPKETAAPTPKPAAPTPKPAAPTPKPAAP 660  
 DB 800 KELAPTTKGPSTTSKAPATTPKETAAPTPKETAAPTPKPAAPTPKPAAPTPKPAAP 859  
 QY 661 PTTTKEPTTIHSSPDESPTBELSAEPTPKALESKPEGVPTTKTPAATKPEMTTAKDKT 720  
 DB 860 PTTTKEPTTIHSSPDESPTBELSAEPTPKALESKPEGVPTTKTPAATKPEMTTAKDKT 919  
 QY 721 TERDRTTPETTTAAPKMTKETATTEKTESKITATTOYTSSTTTOOTTEFFKITTLAKT 780

DB 920 terdlrtlpeltaapxkmtketalttektteskiflatitqvtsttqdtptfittlkt 979  
 QY 781 TLAPKVTTKTKTTTITIMNKPEETAAPKRDRAINSATTPPKPKPAAPKPTSTAKPKPT 840  
 DB 980 TLAPKVTTKTKTTTITIMNKPEETAAPKRDRAINSATTPPKPKPAAPKPTSTAKPKPT 1039  
 QY 841 MPVRKRPKTPTPRKMSTMPBELNPTSRIAEAMLGOTTTPRPNCPNSKLVENPKRSFDAG 900  
 DB 1040 MPVRKRPKTPTPRKMSTMPBELNPTSRIAEAMLGOTTTPRPNCPNSKLVENPKRSFDAG 1099  
 QY 901 AEGETPMILRPVFMPEVTPMDIYLRPNNOGIIINPMLSDETNCNGKPYDGLTTLRN 960  
 DB 1100 AEGETPMILRPVFMPEVTPMDIYLRPNNOGIIINPMLSDETNCNGKPYDGLTTLRN 1159  
 QY 961 GTLVAERGHYFWMLSFSPSPSPARITEVWGLSPIDVTFQNCNCGKTFPFKDSQYMR 1020  
 DB 1160 GTLVAERGHYFWMLSFSPSPSPARITEVWGLSPIDVTFQNCNCGKTFPFKDSQYMR 1219  
 QY 1021 TNDIKAGYKPIRFKFGGLTGQIYVALSTAKYKKNPESVYEFK 1064  
 DB 1220 TNDIKAGYKPIRFKFGGLTGQIYVALSTAKYKKNPESVYEFK 1263

## RESULT 2

AAAR26049 ID AAR26049 standard; Protein: 1404 AA.

AC AAR26049;

DT 02-FEB-1993 (first entry)

DE MSF precursor.

KW Megkaryocyte colony stimulating factor; secretion signal; meg-CSF; stability; proteolytic cleavage; adhesion; alternative splicing.

OS Synthetic.

FT Key Location/Qualifiers

FT 1..26 /label= Exon\_I  
 FT 26..67 /label= Exon\_II  
 FT 67..107 /label= Exon\_III  
 FT 107..157 /label= Exon\_IV  
 FT 157..200 /label= Exon\_V  
 FT 200..1141 /label= Exon\_VI  
 FT 1141..1166 /label= Exon\_VII  
 FT 1166..1212 /label= Exon\_VIII  
 FT 1212..1266 /label= Exon\_IX  
 FT 1266..1331 /label= Exon\_X  
 FT 1331..1373 /label= Exon\_XI  
 FT 1373..1404 /label= Exon\_XII

PN WO9213075-A.

PD 06-AUG-1992.

PE 17-JAN-1992; 92WO-US00433.

PR 18-JAN-1991; 91US-0643502.

PR 10-SEP-1991; 91US-0757022.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:16 ; Search time 107.17 Seconds  
(without alignments)  
735.411 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1263

Perfect score: 5698  
Sequence: 1 VKDNKKKRTKKKPTPKPPVY.....VAALSTAKYKNPESVYFPK 1064

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID8/gcgdata/geneseq/AA1980.DAT:\*  
2: /SID8/gcgdata/geneseq/AA1981.DAT:\*  
3: /SID8/gcgdata/geneseq/AA1982.DAT:\*  
4: /SID8/gcgdata/geneseq/AA1983.DAT:\*  
5: /SID8/gcgdata/geneseq/AA1984.DAT:\*  
6: /SID8/gcgdata/geneseq/AA1985.DAT:\*  
7: /SID8/gcgdata/geneseq/AA1986.DAT:\*  
8: /SID8/gcgdata/geneseq/AA1987.DAT:\*  
9: /SID8/gcgdata/geneseq/AA1988.DAT:\*  
10: /SID8/gcgdata/geneseq/AA1989.DAT:\*  
11: /SID8/gcgdata/geneseq/AA1990.DAT:\*  
12: /SID8/gcgdata/geneseq/AA1991.DAT:\*  
13: /SID8/gcgdata/geneseq/AA1992.DAT:\*  
14: /SID8/gcgdata/geneseq/AA1993.DAT:\*  
15: /SID8/gcgdata/geneseq/AA1994.DAT:\*  
16: /SID8/gcgdata/geneseq/AA1995.DAT:\*  
17: /SID8/gcgdata/geneseq/AA1996.DAT:\*  
18: /SID8/gcgdata/geneseq/AA1997.DAT:\*  
19: /SID8/gcgdata/geneseq/AA1998.DAT:\*  
20: /SID8/gcgdata/geneseq/AA1999.DAT:\*  
21: /SID8/gcgdata/geneseq/AA2000.DAT:\*  
22: /SID8/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5698	100.0	1299	22	AAAM24322 Human EST encoded
2	5698	100.0	1404	13	AAAR26049 MSF precursor. SY
3	5698	100.0	1404	22	AAAB05658 Human megakaryoct
4	5698	100.0	1404	22	AAAB29773 Human megakaryoct
5	3484	61.1	902	22	AAAB29778 Human MSF-derived
6	1357.5	23.8	452	16	AAAB00041 Human megakaryoct
7	1168	20.5	5179	22	AAAM24516 C899P predicted am
8	1042.5	18.3	472	22	AAAB05659 Bovine MSF ortholo
9	981	17.2	763	21	AAAG38942 Arabidopsis thalia
10	950	16.7	1664	19	AAAW3106 C. thermocellum OI
11	771.5	13.5	1325	22	AAAM03645 Peptide #2327 enco

12	715	12.5	763	18	AAAM31852
13	625.5	11.0	4412	21	AAAY3666
14	612	10.7	572	18	AAAM31855
15	544	9.5	844	7	AAAB05070
16	542	9.5	807	21	AAAY54467
17	520.5	9.1	788	21	AAAY54466
18	508	8.9	1837	21	AAAB11726
19	506.5	8.9	744	9	AAAB29725
20	496.5	8.7	2971	21	AAAB41231
21	489.5	8.6	2972	22	AAAB50363
22	489.5	8.6	3118	22	AAAB50362
23	489	8.6	1721	21	AAAB11727
24	488.5	8.6	826	13	AAAR26042
25	488	8.6	617	22	AAAM16458
26	488	8.6	617	22	AAAM04187
27	488	8.6	957	21	AAAY59288
28	488	8.6	957	22	AAAM24513
29	488	8.6	1721	19	AAWA8299
30	485	8.5	1127	22	AAAB5541
31	476.5	8.4	652	9	AAAB2974
32	467.5	8.2	511	22	AAAM14883
33	467.5	8.2	511	22	AAAM27312
34	467.5	8.2	511	22	AAAM02607
35	450.5	7.9	378	12	AAAR14160
36	446.5	7.8	378	12	AAAR14162
37	446.5	7.8	750	20	AAAY05477
38	443	7.8	2870	21	AAAY5559
39	443	7.8	3178	21	AAAY5556
40	442.5	7.8	751	16	AAAR0839
41	433	7.6	2819	22	AAAB35408
42	428.5	7.5	2665	22	AAAM14533
43	428.5	7.5	2665	22	AAAM26950
44	428.5	7.5	2665	22	AAAM2259
45	428.5	7.5	3266	21	AAAB42491

## ALIGNMENTS

RESULT 1  
ID AAM24322 standard: Protein: 1299 AA.  
AC AAM24322;  
DE 12-OCT-2001 (first entry)  
DT Human EST encoded protein SEQ ID NO: 1847.  
DX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biotechnology; gene therapy; nutrition.  
XX  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
DR WPI; 2001-476164/51.

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the Pa  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precu  
Human ORFX ORF995  
Human SRCAP. Homo  
Human SRCAP. Homo  
Portion of cryptos  
P. yoelii SSP2 ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 poly  
9900P predicted am  
Cryptosporidium pa  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
PRP 378. Triticum  
PRP encoded by clo  
C. albicans Rbt1 p  
Caenorhabditis ele  
Caenorhabditis ele  
Japanese sea musse  
Human 07CG27 gene  
Peptide #967 enco  
Peptide #987 enco  
Peptide #941 enco  
Human ORFX ORF2255

CC combinations of alternatively spliced exons from this sequence, with  
 CC the exons spliced together in differing orders to form different  
 CC members of the MSP family.  
 XX  
 SO Sequence 111 AA;

Query Match 8.3%; Score 613; DB 13; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-26;  
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 LSSCAGRCGEYSRDATCNDYNCQHYWECCPDFKRVCTAELSCGRCFESFERGRECD 62  
 Db 8 LSSCAGRCGEYSRDATCNDYNCQHYWECCPDFKRVCTAELSCGRCFESFERGRECD 67  
 QY 63 DAOCRRYDKCCPDYSPCAEVHNPSPSSKKAPPSGASQTIK 106  
 Db 68 daqckkydkcpcdyfcaevhnpstppsskkappspgaaqtlk 111

Search completed: April 26, 2002, 16:16:03  
 Job time: 154 sec

E-1.4













PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161930.  
PR 28-OCT-1999; 99US-0161992.









|||||  
Db 865 eptlhkspeestpelsaepfklenspkpvgvltktpaatkpemttakkdtkterel 924  
QY 901 RTTPETTTAARKKTKETATTTTEKTSKITATTTQVSTTODTTPKTTTKTTTAK 960  
Db 925 rttepettaapkmktetatektetktskltatltqvtsttdtptkltlktlajk 984  
QY 961 VTTTKTTTITTEIMNKPEETAKPKDRATNSKATTPKQKPTKAKKPTSTKKPKTPRVR 1020  
Db 985 vtttkktlttelmkpeetaktakdratskaltkpkqkptkpkstaktkpktmpvr 1044  
QY 1021 KPKTTPPKKMTSTMBELNPTSRIAPAMQTTPRNQTSKLIENVPKSEDAAGAEGET 1080  
Db 1045 kptktpkpkmtstmbelnpktsriaeamlqtltrpnpknsklivenpksedagaeget 1104  
QY 1081 PHMLARPHFEMPEVTPDMVLPVPNOGIIINPMLSDETINICNGKRVDDGLTTLRNGTLVA 1140  
Db 1105 phmlarphfempevcpdmvlpvpnggiiinpmisdenicngkpvdglttlngtlva 1164  
QY 1141 FRGHYFMMLSPPSPSPARITEWGISPIDVTFRNCNCGKTEFFKOSQIWRFTNDIK 1200  
Db 1165 frghyfwmlspfsparparriletwgispidvtfrncncegkttffkdsqywrftndik 1224  
QY 1201 DAGYPRPIRKGEGGLTGOIYVAALSTAKYKMWPEVYFKRGSIGQYIKQEPVOKCPGR 1260  
Db 1225 dagypkpiRkg19glqglvaalstakymwpevyfkrgs19qy1ykqepvokcpgr 1284  
QY 1261 RPLANPVGEMTOVRRRRFRERAIGPSQHTTIRIQYSPARLAYODKGVLHNEVKSILMR 1320  
Db 1285 rplnlpvgemtovrrrrtereraigsqhtlirigsparlaydgylhnevkasilwr 1344  
QY 1321 GLPNVTSALSLPNIRKPDGYDYAASKDOYVINDPSRTARAITTRSGOTLSKVMYNCP 1380  
Db 1345 glpnvtsalslplnrkpdgydyaiskdqyindvpsrtaraittrsgotlskvmyncp 1404

RESULT 4  
ID AAM24322 standard; Protein; 1299 AA.  
XX AAM24322:  
XX  
XX  
XX 12-OCT-2001 (first entry)  
XX  
XX Human EST encoded protein SEQ ID NO: 1847.  
XX  
XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
XX  
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
XX  
XX diagnostics; forensic test; gene mapping; genetic disorder;  
XX  
XX biodiversity; gene therapy; nutrition.  
XX  
XX Homo sapiens.  
XX  
XX MO200154477-A2.  
XX  
XX PD 02-AUG-2001.  
XX  
XX PF 25-JAN-2001; 2001WO-US02687.  
XX  
XX PR 25-JAN-2000; 2000US-0491404.  
XX  
XX PR 17-JUL-2000; 2000US-0617746.  
XX  
XX PR 03-AUG-2000; 2000US-0631451.  
XX  
XX PR 15-SEP-2000; 2000US-0663870.  
XX  
XX PA (HXSE-) HXSE0 INC.  
XX  
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
XX  
XX Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX WPI; 2001-476164/51.  
XX  
XX DR N-PSDB; AAH98981.  
XX  
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -  
XX  
XX Claim 20; page 1198-1201; 1275pp; English.  
XX  
XX  
CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
XX Sequence 1299 AA;  
SQ  
Query Match 92.2%; Score 6834; DB 22; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QDLSGAGCGSGYSRDANCNDYNOHMECCPDKRYCTAELSGRCFSEFGRREC 60  
Db 25 qdlsscaagrcgsgysrdancndynohmccpdkryctaelsgyrcfsefgrrec 84  
QY 61 DCDAGCKKYDKCCPDYSESCAEVHNFTSPSSSKAKAPPASQTIKSTTKRSPKPNKK 120  
Db 85 dcdagckkydkccpdyse scaevhnftspssskakappasqtlksttkrspkpnkk 144  
QY 121 TKKVIESEITTEHSHVSENOESSSSSSSTTWIKISKNSANRELQKKIWKDKK 180  
Db 145 tkkvieseeitteeshsvenoessssssssttwikisksnsanrelqkkikvdkkk 204  
QY 181 KNRTEKKKPPKPPVVDAGSGLDNGDFKVTTPSTTQHNKYSTSPKITANKPINRPSL 240  
Db 205 knrtekkkppkppvvdagsgldngdfkvttpsttqhnkystspkitankpinrpsl 264  
QY 241 PPNSDTSKETSILTVNKEETVEKETTNNKQISTDGEKETSKEKQSIIEKTSANLAPT 300  
Db 265 ppsndtsketsilvtnkettvekettnnkqistdgeketsakeqsieksaklapt 324  
QY 301 SKVLAKPTPKAETTTGPAITTPKKEPTTPPKKPASTTPKEPTTIKSAPTTPKPAPT 360  
Db 325 skvlakptpkacttgpalttppkkep ttpkpkasttpk epttiksap ttpkpa pt 384  
QY 361 TTKSAPPTPKKPAETTTKEPAETTPKKEPAETTTKKEPAETTTKSAPPTPKKPA 420  
Db 385 ttksap ttpk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk kpa 444  
QY 421 PTPPKKPAETTPKKEPTTPPKKPAETTPKKEPAETTTKKEPAETTTKKEPAETTTK 480  
Db 445 ptpkkp aet ttpk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk 504  
QY 481 EPAPTTKEPSPTPKKPAETTTKSAPTTTKKEPAETTTKSAPTTTKKEPSPTTKKEPA 540  
Db 505 epap ttpk eps ptpk kpa et ttpk sap ttpk kpa et ttpk sap ttpk kpa 564  
QY 541 PKKPAETTPKKEPAETTTKKEPAETTTKKEPAETTTKKEPAETTTKKEPAETTTK 600  
Db 565 pk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk k 624  
QY 601 TPTPKKLAETTPKKAETTPKKAETTPKKAETTPKKAETTPKKAETTPKKAETTPK 660  
Db 625 tptpk kla et ttpk ka et ttpk ka et ttpk ka et ttpk ka et ttpk ka 684  
QY 661 PKKPAETTPKKEPAETTTKKEPAETTTKKEPAETTTKKEPAETTTKKEPAETTTK 720  
Db 685 pk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk kpa et ttpk 744  
QY 721 TSDKPAETTPKGAETTPKKEPAETTTKGAETTPKKAETTPKKAETTPKKAETTPK 780  
Db 745 tsdkp aet ttpk ga et ttpk ke pa et ttpk ka et ttpk ka et ttpk ka 804  
QY 781 TTTGPGTSTSDKPAETTPKKEPAETTTKKEPAETTTKKEPAETTTKKEPAETTTK 840



QY	1081	PHMLLRPHVHEPTEPTPMOKLVPRVPMNGITINPMLSDEFINICNGKRVDCGLTTLRNGTLVA	1140
Db	1105	PMMLILTRPHVIMPEVCPMDMDYLPYRPMQGIILMPMLSDENLGNCPVGDITLITRGTLVA	1164
QY	1141	FRGHVFWMLSPFSPSPSARRITEWGCIPSEIDIVFRMCCEKTEFFFDOSQVWRFTNNIK	1200
Db	1165	FRGHVFWMLSPFSPSPSARRITEWGCIPSEIDIVFRMCCEKTEFFFDOSQVWRFTNNIK	1224
QY	1201	DAGVPRPLFRPGFGGLTQCOIYAAALSTAKYKNWPESEVYFFFRKGSITQOITYKQEPVOKCPGR	1260
Db	1225	DAGVPRPLFRPGFGGLTQCOIYAAALSTAKYKNWPESEVYFFFRKGSITQOITYKQEPVOKCPGR	1284
QY	1261	RPALNYPVUYGEMTOVRRRRRPERAIGSSQHTTIRIQYSPARLVAODKGLVHNEVKSIILMR	1320
Db	1285	RPALNYPVUYGEMTQVRRRRRPERAIGSSQHTTIRIQYSPARLVAODKGLVHNEVKSIILMR	1344
QY	1321	GLPWNWTSALSLNPNRKPDQDYVYAAKSKOQYNIQVPSNSTAAITRTSGQTLRKWVNCIP	1380
Db	1345	GLPWNWTSALSLNPNRKPDQDYVYAAKSKOQYNIQVPSNSTAAITRTSGQTLRKWVNCIP	1404

RESULT	3
AAB29773	
ID	AAB29773 standard; Protein: 1404 AA.

DT 28-FEB-2001 (first entry)

Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.

KW Human MS; megakaryocyte stimulating factor; ribonectin;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KW osteoarthritis; ribosupplementation; tissue adhesion inhibition;  
KW friction coefficient reduction; gene therapy; antiarthritic;  
KW osteopathic.

OS Homo sapiens.

PN WO200064930-A2.

PD 02-NOV-2000.

24-APR-2000; 2000WO-US10953.

PR 23-APR-1999; 99US-0298970.

PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER  
XY

PI Jay GD;  
vy

DR WPI; 2001-024673/03.  
DR N-PCDP; 2001-024673/03.

DR N-PSDB; AAC81498.

novel triboinsectin polypeptide useful as lubricant for treating osteoarthritis, comprises O-linked lubricating moiety -

Claim 3; Page 7; 47pp; English.

the invention relates to a human tribonectin which is a product of an alternative splicing of the human MSF (mesangiocytoc stimulating factor) gene. The tribonectin has at least one O-linked oligosaccharide lubricating motility and has a polypeptide sequence comprising 1-76 repeats of a motif having at least 50% identity to the sequence KEPAPT (AA829774). The invention also relates to a nucleic acid encoding a human MSF-derived tribonectin; a biocompatible composition comprising a human tribonectin for inhibiting tissue adhesion formation; and a method of diagnosing osteoarthritis or a predisposition to osteoarthritis by measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating

CC mammalian joints, such as articulating joints of humans, dogs or horses.  
 CC The tribonecstin, when formulated as a membrane, foam, gel or fibre, is  
 CC useful for inhibiting adhesion between two surfaces such as the injured  
 CC tissues of a mammal, where the injury is caused by a surgical insertion  
 CC or trauma, or an artificial device e.g., an orthopedic implant. In  
 CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
 CC tribonecstin may be used in gene therapy. The present sequence represents  
 CC human MSF.  
 XX  
 SQ Sequence 1404 AA;

Query Match	100.0%	Score 7410;	DB 22;	Length 1404;
Best Local Similarity	100.0%;	Pred. NO. 0;		
Matches 1380;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	ODLSAGACGCGYSGYSDADNCNDYNOQHMECCPDKRCFVCTAELSGKGFSEFEEGRGRC	60
Db	25	qdlsscagrcgsgyrdalcncdncqhyamecpdtkrrvctaelscgkycfcesfegrgrec	84
QY	61	DODDACCXYKDCCPDYESTCAEVHNPSTPSSKKAPPPSGASOTIKSTYKRSBPNNKKR	120
Db	85	cdcaeqckxydkccpdyestcaeivhnpstpskkappsgaaqqlkstkckrppknkk	144
QY	121	TKKVISEBTEBHSVSENOQSSSSSSSSSSSTTIWKIKSSKNNSANRELQKKLJKYDKMK	180
Db	145	tkkviaseeileehsvsenqessssssssstlwkikssknsaanrelqklkkydkmk	204
QY	181	KNRRTKKRPKPPVYVDEAGSGLDNDFKTYTTPDSTYQNNKYSTSKTITYAKPINRPSL	240
Db	205	knrttkkrrppkppvyvdeagsgldnfdktyttpdstyqnnkystsktityakpinrpsl	264
QY	241	PENSDTSKETSITVYKETTETVETKEETTTNNKQSTQGKEKTKTAKENQSIERTKSAKLAPY	300
Db	265	pnpsdtsketsitvnyketvettvettlnkqstidgkektaketsakeqslktsakldlapt	324
QY	301	SKVLAKPPPKAETTKGVALTPKREPTPTTPKREASTTPKERTPTTIKASAPTTKREAPAT	360
Db	325	skvlakpppkaeettkgvaltcpkaptptcpkrepastcpkaptpltkasapttkreapapt	384
QY	361	TTYSAPPTPKREAPPTTKREAPATTKREAPATTKREAPATTKREAPATTKREAPATTKRPA	420
Db	385	ttksapttkreapapttkreapapttkreapapttkreapapttkreapapttkreapapttkrpa	444
QY	421	PTTPEKEAPTTPKERTPTTPKREAPATTKREAPATTKREAPATTKREAPATTKREAPATTK	480
Db	445	pttpkeappttkertpttpkrepapttkreapapttkreapapttkreapapttkreapapttk	504
QY	481	EPAPTTTKESBPTTKREAPATTKSAPTTTKREAPATTKSAPTTTKESBPTTKREAPATTK	540
Db	505	epaptttkesbpttkreapatttksaptttkreapatttksaptttkesbpttkreapatttk	564
QY	541	PKEBAPTTPPKAPATTPKREAPATTKREAPATTKKAPATAKREAPATTKKETAAPTTPKKL	600
Db	565	pkeapapttkkapapttkreapapttkreapatttkkapatakreapatttkketaapttpkkl	624
QY	601	TTTTPEKLAAPTTPEKAPATTPPEELAPTTPEBPTPTTPEBAPATTKPKAAAPNTPKREAPAT	660
Db	625	ttpttpeklaapttpekapaattpeelapttpeebptpttpeebapatttkpkaaapntpkreapatt	684
QY	661	PKREBAPTTPKREAPATTPKREAPATTKPGKAPATTKLEAPATTKKAPAKLAAPTTKEPST	720
Db	685	pkreapapttkreapatttkreapatttkpgkapatttkleapatttkkapaklaapttkepst	744
QY	721	TSDKAPATTPKGTAPTTPKREAPATTKREAPATTKPGTAPTTTLKLEAPATTKPKKAPKELAP	780
Db	745	tsdkapatttkgtaapttpkreapatttkreapatttkpgtaaptttlkleapatttkpkpkkelap	804
QY	781	TTTTGGPSTTSDKAPATTPKETAAPTTKREAPATTKPKKAPATTPPETPTPTTSEVSTPYTTK	840
Db	805	tttkgptsttsdkapaatttketaaptttkreapatttkpkkapattppetptpttsevastpyttk	864
QY	841	EPPTTIHKSPOESTPELSAEPPTPKALENSPKBGPVPTTKTPAATPEMUTTTAKDSTTTERDL	900

OY 1261 RPALNYPYGEWQYRRRRFERALGPSTHTIRIOYSPARLAYODKGVLAHNEVKYSILMR 1320  
DB 1285 rpalnypygemtqyrrrrrrferaisqsthtir:igysparlayqdkgylhnevkysilwr 1344  
OY 1321 GLPNVYSAISLPNIRKFDGIDYVAFSKDQYINIDVPSRTARATITRSGQTLSTKWYMC 1380  
DB 1345 glnpvnysaaislpnirkpdyafyskdylnidvpsrtaratitrsqtlstkwynop 1404

RESULT 2  
AAB60568 standard: Protein: 1404 AA.  
ID AAB60568  
AC AAB60568:  
DE 27-APR-2001 (first entry)  
XX Human megakaryocyte stimulating factor (MSF, CACP).  
XX Human: CACP protein; campitodactilyl-arthropathy-coxa vara-pericarditis;  
XX MSF; megakaryocyte stimulating factor; synovial lubricant;  
XX chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
XX antarthritic.  
XX Homo sapiens.  
XX WO200107068-A1.  
XX 01-FEB-2001.  
XX 21-JUL-2000; 2000WO-US20002.  
XX 23-JUL-1999; 9905-0145328.  
XX 19-JUL-2000; 2000US-0145328.  
XX (UYCA-) UNIV CASE WESTERN RESERVE.  
XX Warman ML;  
XX WPI: 2001-182721/18.  
XX New composition comprising the campitodactilyl-arthropathy-coxa  
XX vara-pericarditis protein in combination with an anesthetic, useful for  
XX treating osteoarthritis, or as lubricants of tissue and joints  
XX Example 1; Page -: 34pp; English.  
XX The invention relates to a method of treating osteoarthritis via the  
XX administration of a composition comprising the campitodactilyl-arthropathy-  
XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
XX The composition may further comprise a local anesthetic. The composition  
XX of the invention may be administered via intra-articular or intravenous  
XX injection. The human CACP protein is identified in the invention as  
XX being megakaryocyte stimulating factor (MSF), and mutations in  
XX CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
XX this gene are responsible for the heritable disorder campitodactilyl-  
XX arthropathy-coxa vara-pericarditis, in which patients have synovial  
XX hyperplasia without evidence of inflammation. CACP protein (MSF)  
XX acts as a synovium lubricant, and can be used to lubricate tissue and  
XX joints in the treatment of osteoarthritis. The composition may be  
XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
XX loss of range of movement or joint damage). The present sequence  
XX represents human megakaryocyte stimulating factor (MSF, CACP protein).  
XX Note: This sequence is not given in its entirety in figure 4 of the  
XX specification, although a Genbank accession number was given. This  
XX sequence was therefore obtained from Genbank (U70316).

Sequence 1404 AA:  
Query Match 100.0%; Score 7410; DB 22; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QDLSSACRCEGEGYSRDATCNQDINQHYMCPCDFKRVCAELSCGRCEPSEFREG 60  
DB 25 qlssacagrcgegysrdatcncdqncqhymcpcdfkrvcaelscgrcfesfeirec 84  
OY 61 DCDQCKKYDCPCPDYEFCAEHNPTSPSSKAPAPPSGASQTIKSTTKRSPKPKK 120  
DB 85 dcdqckkydcpdpdyefcaeahnptspsskappaappsgasqtlksttkrspkpkk 144  
OY 121 TKVIESEITEEHSVSENOSSSSSSSSSSSTIWKTKSSKNSAANRELQKILKVDNK 180  
DB 145 tkvieeseiteehsvseengssssssstlwkikssksnaanrelqkilkvdnk 204  
OY 181 KNRTKKKPPKPPVVDGSGIUNGDFKVTTPDSTGHNKVSPTTKTAKPINPRSL 240  
DB 205 knrtkkkppkppvvdgsgidngdfkvtppdustcgmnyvspttktkpinprsl 264  
OY 241 PPSDTSKETSILTVNKETVETKETTNNKOTSDGKEKTSAKETOSIEKTSAKDLAPT 300  
DB 265 ppsdtsketsiltnketvelkettnkqtsldgketsksetgsiektsakdlapt 324  
OY 301 SKVLAKPYPKATTTKGAALTTPKREPPTTPKREASTPKREPPTTIKSAPTTPEAP 360  
DB 325 skvlpkypkaettkgpalttkpkepdttkpkeasttkpkepttklsapdtkeapt 384  
OY 361 TTKSAPTPPKREPAPTTKREPAPTTKREPAPTTKREPAPTTKSAPTTKREPAP 420  
DB 385 ttksaptkpkeapttkkeapttkkeapttkkeapttkkeapttkkeapttkkpa 444  
OY 421 PTPPKREPAPTTKREPPTTPPKREPAPTTKREPAPTTKREPAPTTKREPAP 480  
DB 445 ptkpkeapttkpkepttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapt 504  
OY 481 EPAPTTKKEPSTTPKREPAPTTKSAPTTKREPAPTTKSAPTTKKEPSTTPKREPAP 540  
DB 505 epapttkkespttkrepapttksapttkrepapttksapttkkesapttkrepap 564  
OY 541 PKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAP 600  
DB 565 pkeappttkpkeapttkpkeapttkpkeapttkkpaapkeapttkpkeapttkk 624  
OY 601 TPTPKKLAPTTPKREPAPTTPEELATTPPEEPPTTPPEEPAPTTKAAAPNTPKREPAP 660  
DB 625 tptpkklapttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapt 684  
OY 661 PKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAP 720  
DB 685 pkeappttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapt 744  
OY 721 TSDKPAPTTPKGTAPTTKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAP 780  
DB 745 tsdkpapttkgtapttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapt 804  
OY 781 TTKKGSTSTSDKPAPTTPKREPAPTTKREPAPTTKREPAPTTKREPAPTTKREPAP 840  
DB 805 ttkkgststsdkpaapttkpkeapttkpkeapttkpkeapttkpkeapttkpkeapt 864  
OY 841 EPTTIKSPDESPBELSAEPTPKALENSPKPEGVPTTKTPATKPMPTTAKDKTTER 900  
DB 865 epttikspdespelsaeptpkalenspkpgevppttktpatkpmpttakkdktte 924  
OY 901 RTTPETTTAAPKMTKEATTTETTESKITATTTQVTSNTTQOTTPFKITTLTKTTLA 960  
DB 925 rttpetttaaapkmtkeatttetsteskitatvtstqtdtptfkittlkttlap 984  
OY 961 VTTTKKTTTTELMNKPPEEPAKPKDRATNSKATTPKOKTAPKAPKPTSTKPKTMP 1020  
DB 985 vtltkkttttelmnkpeepakpkdratnskattpkpkapkpstkpktmpvr 1044  
OY 1021 KPKTTPPKRMTSMPELNDTSRIAEAMLOTTTPNQTNSKILVEYNPKSEADGAGBGT 1080  
DB 1045 kpkttpprkmtsmpeelnptsriaeamlottpnqtnskilveynpkseadagagbgt 1104





1-3



```

XX 11-SEP-1984; 84AU-0007067.
PR 11-SEP-1984; 84AU-0007066.
PR 10-SEP-1985; 85AU-0047326.
XX
PA (HALL-) HALL INST MED RES.
XX
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR MPI: 1986-094065/14.
DR N-PSDB: AAN60473.
XX
PT DNA coding for plasmodium falciparum antigens - expressing
PT polypeptide(s) having antigenicity of RESA or FIRA antigens of P
XX falciparum
XX
PS Disclosure; Fig 7; 55pp; English.
XX
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (AAN60472) or FIRA (AAN60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC plasmodium falciparum malarial infections.
XX
SQ Sequence 844 AA:

```

```

Query Match 7.6%; Score 555.5; DB 7; Length 844;
Best Local Similarity 25.7%; Pred. No. 2.8e-22;
Matches 252; Conservative 128; Mismatches 358; Indels 241; Gaps 49;

```

```

OY 111 IESEETHEHSYSENOESS-SSSSSSSSSTIMKIKSSKKSANRELQKKLYKDNKN 168
Db 1 mesgkkaaspsinvdeyslsnnepnqatlnltpdqslnahlnneinle----- 55
OY 169 RTKKRTPRPVVDAGSGLDNGDFKVTTPDSTTOHKNKYSTSPKITAKPINRPSLP 228
Db 56 -----tlstlttnenevnpivpsis-nlclcltp 86
OY 229 NS-----DTSKETSIVNKETVETKETTNNKOTSTDGKEKTTSAKETOSIEKTS 279
Db 87 yqlililislvnfcckkqlynk--nfeekfnlasvgsnatqgensnqknelev-kes 143
OY 280 AKDLAPTSTKVLAKPTPKAETTTKGPALTPKEPT-----PTPKKEPASTPKKEPTTIKS 335
Db 144 sqtqpv-----tpqetvtlq--aaatpgetvteqevltleevlt--qevlt--q 189
OY 336 APTPKKEPAPTTTSAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 391
Db 190 epltvqep--vltvqevltvqep--pvtvqep--pvtvqepvltvqevltvqep--vtsqep 239
OY 392 TTPKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTP 447
Db 240 vtpqe-----vltvqep--pvtvqep--pvtvqepvltvqevltleevlt 284
OY 448 APKRPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 499
Db 285 -----qevltle--pvtvqepvltleevltvqevltvqevltvqep--vltvqevltvqep 335
OY 500 APT-----TKSAPTTPKESPTTKEPA-----PTTKEPA 531
Db 336 vtveehidekkqsegdnlsissiseeteeekshlkkkswlkfgnknkdkknekkps 395
OY 532 PTPPKRPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTP 591
Db 396 leavkqnadeqkqpcdsqsvnagdvltqdeplacqep-pltqelta----- 442
OY 592 EKLAFTTPEKAPTTPEELAPTTPEEBPTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPE 647
Db 443 -----tqe--pltqep--vltvqepvltqep--vltvqepvltvqepvltvqepvlt 485
OY 648 KE-----PAPTTTKEPAPTTTTPKETAFTTTPKETAFTTTPKETAFTTTPKETAFTT 703
Db -----tqe--pltqep--vltvqepvltqep--vltvqepvltvqepvltvqepvlt 703

```

```

Db 486 qeliatqepstqehadekkasegdnlsislseeteeekshlkkkswlkfgnknk 545
OY 704 TSTSDKRAPPTPKCTATPTTKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAP 763
Db 546 ksknekkpslesvkqnadeqkqpcdsqsvnagdvltqdeplacqep-pltqelta 604
OY 764 LAPTTTQGPSTTSDKRAPPTTTPKKEPAPTTTTPKKEPAPTTTTPKKEPAPTTTTP 823
Db 605 -pltq-----vltvqepvltqep--vltvqepvltqep--vltvqepvltvqepvlt 648
OY 824 TTKETPTTHKSPDE-----STPELSAEPTRKA-----LENS 854
Db 649 atqepstqehadekkasegdnlsislseeteeekshlkkkswlkfgnknkdkk 708
OY 855 PKEPQVPTTKTPAATKPEMTTAKDXTTERDLRTTPE-TTAAFKMTKEATTEKTTES 913
Db 709 ekpslesvkqnadeqkqpcdsqsvnagdvltqdeplacqep-pltqelta 768
OY 914 KITATTTQVSTTTQODTTTFR--ITTLKTITTLAKVTT-----TKTITTTETIMNKEET 966
Db 769 e-----vltvqepvltqepvltqepvltqepvltqepvltqepvltqepvltqepvlt 823
OY 967 AKPKDRATNSKATTPPKPOK 985
Db 824 asegdnlsislseeteeek 842

```

```

Search completed: April 26, 2002, 16:25:13
Job time: 503 sec

```

























DB 721 apctpkpapkclaplttkcsttsdckpaptkpglapltkcapltkpkapltkpg 780  
QY 743 TAPPTLKPPAPTPPKPAPKAPKELAPTTTKGPTSTSDKAPATTPPKAPTPPKAPTPPK 802  
DB 781 taptltkcapltkpkpapkclaplttkcsttsdckpaptkpglapltkcapltkpkapltk 840  
QY 803 KPAPTPEPTPPPTSEVSTPTTKEPTTIKSPDESTPELSAEPTRKALENSKPECPVT 862  
DB 841 kpatltkcapltkpkpapkclaplttkcsttsdckpaptkpglapltkcapltkpkapltk 900  
QY 863 TKTPPAKPKPMPTTAKKKTTERDLRTPEPTTAPAKMTKATATTEKTSKIRATTOY 922  
DB 901 tktpaakpkemtakckltterdlrtpepttapaakmtkattkctkctkctkctkctkctk 960  
QY 923 TSTTTODTTPPKITTKTTLTTLAPKVTTKTKTITTEITMKNPEETAKPKDRATNSKATTPK 982  
DB 961 tsstttodttpfkittlttlttlapkvtttkittlttelnmkpeetackkdratnskattpk 1020  
QY 983 POKPTKAPKPTSTKTKKTMKPRVAKKPTTPPKMTSTMEPLNPTSKIAEAMLOTTTRPN 1042  
DB 1021 pdkptkpkpkstckkpktmpvrvkpkltcpkmtstmpelnptskiaeamlqtltrpn 1080  
QY 1043 QTPMSKLVENPKSEDGAGSETPHMLLRPHVFMPEVTPMDVLPVPNOGIINPMLS 1102  
DB 1081 qtpmsklvenpkseadaggaetphmllrphvfmpevtmdvlpvpnggillnplms 1140  
QY 1103 DETNICKGKPYDGLTTLRNGTLVAFRGHYFMWMLSPSPSPARRITTEVWGISPTDVT 1162  
DB 1141 detnckgkpydgltlrngtlvafrghyfmwmlspfsparritlewgsipstidvft 1200  
QY 1163 RCNCEGTFPPKDSQVRFNTDIDAGVPRKFEKPGSLTGOIYAAISTAKYKMPESVY 1222  
DB 1201 rcncegtffikdsgvrfndidagvprkfkfglqgivaastakaykmpesvy 1260  
QY 1223 FFKRGSGIQOITYKQEPVOKCPGRRPALNYPVYGEMLQVRRRRFERAIGPSQHTTIRIOY 1282  
DB 1261 ffrkrgsgiqiykqepvqkcpgrrrpalnypvygemlqvrrrrferaisgsqhtlirioy 1320  
QY 1283 SPARLAVODKGVLEHNEKVSILMRGLPNVNTSAISLPNIRKPDGYDYAASKDOYINIDV 1342  
DB 1321 sparlaydkgvlehnkvslmrglpnvntsaistlpnirkpdydyafaskdgyinidv 1380  
QY 1343 PSRTARAITTRSGOTLSKWNCP 1366  
DB 1381 psrtaraaitrsgtlskwnpc 1404

RESULT 4

AAM24322 standard; Protein: 1299 AA.  
ID AAM24322;  
AC AAM24322;  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST encoded protein SEQ ID NO: 1847.  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
OS Homo sapiens.  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-USO2687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSEO INC.  
XX  
XX Tang YT, Lau C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX WPI, 2001-476164/51.  
DR N-PSDB; AAH98981.  
XX  
PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
PS  
PS Claim 20; Page 1198-1201; 1275pp; English.  
XX  
XX  
CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
SQ Sequence 1299 AA;

Query Match 91.7%; Score 6689; DB 22; Length 1299;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1257; Conservative 0; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAMKTPYLLLLLSVFIQOVSSODLSSCAGRCGEYSRPATNCPCYOHMECCPDF 60  
DB 1 mawktlpyllllsvfiqovssodlsscagrcgeysrpatncpcyohmccpdf 60  
QY 61 KRVCTA-----AEVHNPTSPSSKRAP 82  
DB 61 krvtctaelscgrcfesfergedcdagckkykccpdyesfaevhnpspskrap 120  
QY 83 PPSGASQTIKSTYKRSKPKPKKTKKYIESEETIEHVSSENOSSSSSSSSSTIWM 142  
DB 121 ppsgasqtkstkrspkpkktkkyieeseetiehvseneqssssssssstliw 180  
QY 143 KIKSSKNSAANRELQKLVKDNKNKRTKPKRPVVDAGSLDGDGVKVTTPDST 202  
DB 181 kikssksaanrelqklvxdnknrtkpkrpvvdagsgldgdgkvlttptdst 240  
QY 203 TOHNKVSSTPKITTAAPINRPSLPPNSDTSKETSLIVNKEITYETKETTNNKOTSDG 262  
DB 241 tqhnkvstspkittakpinnrpslppnsdstsketslvnketvettcttnkqtsldg 300  
QY 263 KEKTTSAKETOSIEKTSKADLAFTSKVLAKPTPKAETTTKGPALTTKREPPTTPPKEPAS 322  
DB 301 kektsaketosiektsakdlaptskvlakptpkraetttkgrpalttkreppttpkrepas 360  
QY 323 TTPKEPTTITKSAVTPPKKEPAPTTTKSAPTPPKKEPAPTTTKKEPAPTTTKKEPAPTTKEP 382  
DB 361 ttpkepttltksavtpkpkkepapttkksaptpkpkkepapttkkepapttkkepapttkkep 420  
QY 383 APPTTKSAPTPPKKEPAPTTPKKPAVTPPKKEPAPTTPKPEPTTPPKKEPAPTTKPA 442  
DB 421 apttksaptpkpkkepapttkpkpavtpkpkkepapttkpepttpkpkkepapttkpkp 480  
QY 443 EPAPTPAKKPAVTPPKKEPAPTTTKKESPTTPPKKEPAPTTTTSAPTTTKKEPAPT 502  
DB 481 epaptpakpkpavtpkpkkepapttkkespttpkpkkepapttkksaplttkkepapt 540  
QY 503 TTKSAPTPKKEPSPTTKKEPAPTTPKKKEPAPTTPKKAPATTPPKKEPAPTTPKK 562  
DB 541 tksaptpkkespttkkepapttkpkkepapttkpkkepapttkpkkepapttkpkp 600  
QY 563 APAPKKEPAPTTPKKETATPTPKKILPTTPEKLAPTTPKKEPAPTTPEELAPTTPEEPPTT 622

QY 983 PQRTPAKPKPTSTKKRKTMRVRKPKTTPPKMTSTMPPLNPTSRIAEMLOTTTRPN 1042  
DB 1021 PQRTPAKPKPTSTKKRKTMRVRKPKTTPPKMTSTMPPLNPTSRIAEMLOTTTRPN 1080  
QY 1043 QTPRSKLVENVKSEDAAGAGERTPHMLLRPHVMPPEVTPMDVLPVYVNGIINPMLS 1102  
DB 1081 QTPRSKLVENVKSEDAAGAGERTPHMLLRPHVMPPEVTPMDVLPVYVNGIINPMLS 1140  
QY 1103 DETNICKGKPVQGLTTLRNGTLVAFRGHYEWMMLSPFSPSPARITTEWGIIPSPIDVFT 1162  
DB 1141 DETNICKGKPVQGLTTLRNGTLVAFRGHYEWMMLSPFSPSPARITTEWGIIPSPIDVFT 1200  
QY 1163 RCNCEGRTFFFKDSQYKRFRTNDIDAGYKPKFFKSGGLGQIYALSTAKYKMPSSVY 1222  
DB 1201 RCNCEGRTFFFKDSQYKRFRTNDIDAGYKPKFFKSGGLGQIYALSTAKYKMPSSVY 1260  
QY 1223 FFRKGSIOQYIKQEPVQKCPGRPALNVYGEVMTQVRRRREPRATIGPQOTTRIRQY 1282  
DB 1261 FFRKGSIOQYIKQEPVQKCPGRPALNVYGEVMTQVRRRREPRATIGPQOTTRIRQY 1320  
QY 1283 SPARLAYQDQGVHLNKKVSLWMLGRLPNVYTSALSLPNIRKPDGYDYAFSKQDYNDIV 1342  
DB 1321 SPARLAYQDQGVHLNKKVSLWMLGRLPNVYTSALSLPNIRKPDGYDYAFSKQDYNDIV 1380  
QY 1343 PSRTARATITRSQGLTSKVNWCNCP 1366  
DB 1381 PSRTARATITRSQGLTSKVNWCNCP 1404  
RESULT 3  
AAB29773  
ID AAB29773 standard; Protein: 1404 AA.  
AC AAB29773:  
XX 28-FEB-2001 (first entry)  
DE Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.  
XX  
XX Human MSF: megakaryocyte stimulating factor; tribonectin;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KM osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KM friction coefficient reduction; gene therapy; antiarthritic;  
KM osteopathic.  
XX Homo sapiens.  
OS  
XX  
XX MO200064930-A2.  
PN  
XX  
XX 02-NOV-2000.  
PD  
XX  
XX 24-APR-2000; 2000MO-US10953.  
PE  
XX  
XX 23-APR-1999; 99US-0298970.  
PR  
XX  
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
PA  
XX  
XX Jay GD;  
PI  
XX  
XX WPI: 2001-024673/03.  
DR N-PSDB; AAC81498.  
DR  
XX  
XX Novel tribonectin polypeptide useful as lubricant for treating  
PT osteoarthritis, comprises O-linked lubricating moiety -  
XX  
XX  
XX Claim 3; Page 7; 47pp; English.  
PS  
XX  
XX The invention relates to a human tribonectin which is a product of  
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonectin has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76  
CC repeats of a motif having at least 50% identity to the sequence KRPAPT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a

CC human MSF-derived tribonectin: a biocompatible composition comprising a  
CC human tribonectin for inhibiting tissue adhesion formation; and a method  
CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by  
CC measuring the amount of MSF or its fragment in a biological sample of a  
CC mammal, wherein an increased amount of MSF compared to a control  
CC indicates the presence of or predisposition to developing  
CC osteoarthritis. The tribonectin and DNA encoding it are useful in the  
CC treatment of osteoarthritis, where they may be used for lubricating  
CC mammalian joints, such as articulating joints of humans, dogs or horses.  
CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is  
CC useful for inhibiting adhesion between two surfaces such as the injured  
CC tissues of a mammal, where the injury is caused by a surgical insertion  
CC or trauma, or an artificial device e.g., an orthopaedic implant. In  
CC particular, one of the surfaces is pericardial tissue. DNA encoding a  
CC tribonectin may be used in gene therapy. The present sequence represents  
CC human MSF.  
CC  
XX  
SQ Sequence 1404 AA:  
Query Match 99.6%; Score 7265; DB 22; Length 1404;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 1366; Conservative 0; Mismatches 0; Indels 38; Gaps 1;  
QY 1 MAWTLPIYLLILSVYIQVSSQDISSCAGRCGEYSRDATNCQYHMECCPDF 60  
DB 1 MAWTLPIYLLILSVYIQVSSQDISSCAGRCGEYSRDATNCQYHMECCPDF 60  
QY 61 KRVCTA-----AEVHNPTSPSSKKAP 82  
DB 61 KRVCTA-----AEVHNPTSPSSKKAP 82  
QY 83 PPSGASQTIKSTKRSKPPNKKTKRVIESEETEEHSVSENOSSSSSSSSSTI 142  
DB 83 PPSGASQTIKSTKRSKPPNKKTKRVIESEETEEHSVSENOSSSSSSSSSTI 142  
QY 121 PPSGASQTIKSTKRSKPPNKKTKRVIESEETEEHSVSENOSSSSSSSSSTI 180  
DB 121 PPSGASQTIKSTKRSKPPNKKTKRVIESEETEEHSVSENOSSSSSSSSSTI 180  
QY 143 KIKSSKSNANRELQKRLKVDKNNKTKRPVYVDEAGSLDNGDFKVTTPRST 202  
DB 143 KIKSSKSNANRELQKRLKVDKNNKTKRPVYVDEAGSLDNGDFKVTTPRST 202  
QY 181 KIKSSKSNANRELQKRLKVDKNNKTKRPVYVDEAGSLDNGDFKVTTPRST 240  
DB 181 KIKSSKSNANRELQKRLKVDKNNKTKRPVYVDEAGSLDNGDFKVTTPRST 240  
QY 203 TQHNKYSTPRTAKINPPSLPPNSDTSKESLVNKKETVETETTKNTKQSTMDG 262  
DB 203 TQHNKYSTPRTAKINPPSLPPNSDTSKESLVNKKETVETETTKNTKQSTMDG 262  
QY 241 TQHNKYSTPRTAKINPPSLPPNSDTSKESLVNKKETVETETTKNTKQSTMDG 300  
DB 241 TQHNKYSTPRTAKINPPSLPPNSDTSKESLVNKKETVETETTKNTKQSTMDG 300  
QY 263 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGALTTTKEPTTPPKSPAS 322  
DB 263 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGALTTTKEPTTPPKSPAS 322  
QY 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGALTTTKEPTTPPKSPAS 360  
DB 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGALTTTKEPTTPPKSPAS 360  
QY 323 TTPKEPTPTTIKSAPTTPKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKEP 382  
DB 323 TTPKEPTPTTIKSAPTTPKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKEP 382  
QY 361 TTPKEPTPTTIKSAPTTPKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKEP 420  
DB 361 TTPKEPTPTTIKSAPTTPKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKEP 420  
QY 383 APTTKSAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPK 442  
DB 383 APTTKSAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPK 442  
QY 421 APTTKSAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPK 480  
DB 421 APTTKSAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPK 480  
QY 443 EPAPTAKKKAPPTPKAPPTTPKEAPPTTKBESPTTPKEAPPTTKSAPTTKKEAPPT 502  
DB 443 EPAPTAKKKAPPTPKAPPTTPKEAPPTTKBESPTTPKEAPPTTKSAPTTKKEAPPT 502  
QY 481 EPAPTAKKKAPPTPKAPPTTPKEAPPTTKBESPTTPKEAPPTTKSAPTTKKEAPPT 540  
DB 481 EPAPTAKKKAPPTPKAPPTTPKEAPPTTKBESPTTPKEAPPTTKSAPTTKKEAPPT 540  
QY 503 TTKSAPPTPKBESPTTPKEAPPTTKBESPTTPKEAPPTTKBESPTTPKEAPPT 562  
DB 503 TTKSAPPTPKBESPTTPKEAPPTTKBESPTTPKEAPPTTKBESPTTPKEAPPT 562  
QY 541 TTKSAPPTPKBESPTTPKEAPPTTKBESPTTPKEAPPTTKBESPTTPKEAPPT 600  
DB 541 TTKSAPPTPKBESPTTPKEAPPTTKBESPTTPKEAPPTTKBESPTTPKEAPPT 600  
QY 563 APTPAKKEAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTP 622  
DB 563 APTPAKKEAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTP 622  
QY 601 APTPAKKEAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTP 660  
DB 601 APTPAKKEAPPTPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTPPKAPPTTP 660  
QY 623 PEEPAPTTPKAAANPTKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 682  
DB 623 PEEPAPTTPKAAANPTKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 682  
QY 661 PEEPAPTTPKAAANPTKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 720  
DB 661 PEEPAPTTPKAAANPTKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPT 720  
QY 683 APTTPKAPKAPKELAPTTTKEPTSTSDKPAVTPPKGTAPTTPKAPPTTPKEAPPT 742



FT Region 1266..1331  
FT /label= Exon\_X  
FT 1331..1373  
FT /label= Exon\_XI  
FT 1373..1404  
FT /label= Exon\_XII  
PN W09213075-A.  
XX 06-AUG-1992.  
PD  
XX 17-JAN-1992; 92MO-US00433.  
PE  
XX 18-JAN-1991; 91US-0643502.  
PR 10-SEP-1991; 91US-0757022.  
XX  
XX (GEMV ) GENETICS INST INC.  
PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
DR MPI; 1992-284660/34.  
DR N-PSDB; AAQ27223.  
XX  
XX New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
PT bacterial and viral infections, etc.  
XX  
XX Claim 1, 2 and 3; Fig 1; 87p; English.  
XX  
XX The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine proceeding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNAs each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, ie. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organisation of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterised by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.  
XX  
XX Sequence 1404 AA:

Query Match 99.6%; Score 7265; DB 13; Length 1404;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 1366; Conservative 0; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAMKTLPIYLILLSVFIYIQOVSSQDISCAGRCGEYSRDATNCNDYNOHYMECCPDF 60  
DB 1 mawktlpiylililsvfivlgvssqdisacagrcgeysrdatncndynqhymeccpdf 60  
QY 61 KRVCTA-----AEVHNTPSPSSKKAP 82  
DB 61 krvctaelsckgrcfesferecdacgckkydkccpdyesifaevhnptpspskkap 120  
QY 83 PPSGASQTIKSTKRSPKPNKKTKKVISEETIEEHSVSENOESSSSSSSSSTTW 142  
DB 121 ppsgaqgltkstktrspkppnkktkkviseetieehsvsenoessssssssstllw 180

QY 143 KIKSSKNSANBELQKLVKDNKKRRTKRPKPVVDEAGSLDNGDKVYTPPTST 202  
DB 181 kiksksnsanbelqklkvdknkkrrtkrpkpvvdeagsldngdkvltptdst 240  
QY 203 TQHNKVSIPKITYTAKPINRPSLPPNSDTSKETSITVNEKETTVEETKNTKOTSDG 262  
DB 241 tqhnkvsipklttakpinrpslppnsdtsketsitvneketvettkntkotsdgs 300  
QY 263 KEKTSAKETOSIEKTSADLAPTSVAKPRAKETTCKGPAITTEKEPTTPKEBAS 322  
DB 301 kektsaketosiektsadlaptsvakpraekettckgpaittekepttpkebas 360  
QY 323 TTPKEPTPTIKSAPPTPKREBAPTTPKSAPTTPKREBAPTTPKREBAPTTPKEP 382  
DB 361 ttpkeptptiksapptpkrebattpkpsapttpkrebattpkrebattpkep 420  
QY 383 APTTKSAPPTPKREBAPTTPKKAAPTTPKREBAPTTPKREBAPTTPKREBAPTTPK 442  
DB 421 apttksapptpkrebattpkkaapttpkrebattpkrebattpkrebattpk 480  
QY 443 EPAPTAPKKPAPPTPKREBAPTTPKREBAPTTPKREBAPTTPKREBAPTTPKREBAPT 502  
DB 481 epaptapkkpaptpkrebattpkrebattpkrebattpkrebattpkrebattpk 540  
QY 503 TTKSAPPTPKREBAPTTPKREBAPTTPKKAAPTTPKREBAPTTPKREBAPTTPKREBAPT 562  
DB 541 tksapptpkrebattpkrebattpkrebattpkrebattpkrebattpkrebattpk 600  
QY 563 APAPKREBAPTTPKREBAPTTPKKTPTTPEKLAATPEKAPATTPPELAATPEEPTPT 622  
DB 601 apapkrebattpkrebattpkktpttpeklaatpekapattppeelaatpeeptpt 660  
QY 623 PEEBAPTTPKKAAPTTPKREBAPTTPKREBAPTTPKREBAPTTPKREBAPTTPKREBAPT 682  
DB 661 peebapttpkkaapttpkrebattpkrebattpkrebattpkrebattpkrebattpk 720  
QY 683 APTPKKPAKRELAATPTTKKEPTSTSDKPAATPKGAPATTPKREBAPTTPKREBAPTTPK 742  
DB 721 aptpkkpaakrelaatpttkkeptstsdkpaatpkgapattpkrebattpkrebattpk 780  
QY 743 TAPPTLKREBAPTTPKKAAPTTPKKTSTSDKPAATTPKREBAPTTPKREBAPTTPK 802  
DB 781 tapptlkrebattpkkaapttpkktstsdkpaattpkrebattpkrebattpkrebattpk 840  
QY 803 KPAPTPEPTPTTPSEVPTTTPKKEPTTHKSDESTPELSAPPTKALENSKEPVGPT 862  
DB 841 kpaptpeptpttpsevppttpkkeptthksdestpelsapptkalenskepvgpt 900  
QY 863 TKPRAATKPEMTTJAKDKTTERDLRTTPTTAAKMTKEIATTTTKEKTESKITATTVQY 922  
DB 901 tkpraatkpemtttakdkttterdlrttptttaakmtkeiattttkeketeskitatvtv 960  
QY 923 TSTTQDTPPEKTLTTLKTTTAKVYTTTKTITTEIINKPEETAKPKDRATNSKATTPK 982  
DB 961 tsttqdtppetkltltlktttakvtytttkttittetiinkpeetapkdratnskattpk 1020  
QY 983 POKPTAPKRPKSTTKKPKTTPRVRKPKTTPPRKMTSTWPELNPSTRIAEALQTTTPRN 1042  
DB 1021 pqpptkapkpksttkpkpmpvrkpktpprkmtstwpelnpstriaeaalqtttprn 1080  
QY 1043 OPMNSKLIVEVANKSEBAGABEETPHMLLRPHVMEVTPDMDYLRPVNOGIIINPMLS 1102  
DB 1081 qpmsklivevanksebagabeetphmlrphvmevtpdmaylrpvnoგიიინპმლს 1140  
QY 1103 DETNINCKRPVGLTTLNRGTLVAFRGHYFMMWLSFSPSPSPARKRTTEWGLSPIDVYFT 1162  
DB 1141 detnincrgpvglttlnrgtlvafrghyfmmwlsfspspsparkrttewglspidvft 1200  
QY 1163 RNCBKGKTEFFEDOSQYWRFTNDIKDAGYKPKPIFKGFGGLTGOIYVALSTAYKKMMPESVY 1222  
DB 1201 rncbkgkttffedosqywrftndikdagypkpkpifkgfggltgoyvalstaykkmpesvy 1260

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:24:52 ; Search time 138.84 Seconds  
(without alignments)  
728.782 Million cell updates/sec

Title: Aa6  
Perfect score: 7294  
Sequence: 1 AAMKTLPLYLILLSVFYIQ.....ARAITRSQGLSKWYNCP 1366

Scoring table: BLOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.Genesec\_1101:\*  
2: /SID88/gcgdata/genesec/genesecp/AA1980.DAT:\*  
3: /SID88/gcgdata/genesec/genesecp/AA1981.DAT:\*  
4: /SID88/gcgdata/genesec/genesecp/AA1982.DAT:\*  
5: /SID88/gcgdata/genesec/genesecp/AA1983.DAT:\*  
6: /SID88/gcgdata/genesec/genesecp/AA1984.DAT:\*  
7: /SID88/gcgdata/genesec/genesecp/AA1985.DAT:\*  
8: /SID88/gcgdata/genesec/genesecp/AA1986.DAT:\*  
9: /SID88/gcgdata/genesec/genesecp/AA1987.DAT:\*  
10: /SID88/gcgdata/genesec/genesecp/AA1988.DAT:\*  
11: /SID88/gcgdata/genesec/genesecp/AA1989.DAT:\*  
12: /SID88/gcgdata/genesec/genesecp/AA1990.DAT:\*  
13: /SID88/gcgdata/genesec/genesecp/AA1991.DAT:\*  
14: /SID88/gcgdata/genesec/genesecp/AA1992.DAT:\*  
15: /SID88/gcgdata/genesec/genesecp/AA1993.DAT:\*  
16: /SID88/gcgdata/genesec/genesecp/AA1994.DAT:\*  
17: /SID88/gcgdata/genesec/genesecp/AA1995.DAT:\*  
18: /SID88/gcgdata/genesec/genesecp/AA1996.DAT:\*  
19: /SID88/gcgdata/genesec/genesecp/AA1997.DAT:\*  
20: /SID88/gcgdata/genesec/genesecp/AA1998.DAT:\*  
21: /SID88/gcgdata/genesec/genesecp/AA1999.DAT:\*  
22: /SID88/gcgdata/genesec/genesecp/AA2000.DAT:\*  
23: /SID88/gcgdata/genesec/genesecp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7265	99.6	1404	13	MSF precursor. Sy
2	7265	99.6	1404	22	Human megakaryocyte
3	7265	99.6	1404	22	Human megakaryocyte
4	6689	91.7	1299	22	Human EST encoded
5	3484	47.8	902	22	Human MSF-derived
6	1936	26.5	452	16	Human megakaryocyte
7	1707.5	23.4	472	22	Bovine MSF ortholo
8	1171	16.1	5179	22	C899P predicted am
9	981	13.4	763	21	Arabidopsis thalia
10	950	13.0	1664	19	C. thermocellum OI
11	802	11.0	1325	22	Peptide #2327 enco

12	715	9.8	763	18	AAM31852
13	640	8.8	4412	21	AAV33666
14	612	8.4	572	18	AAW31855
15	555.5	7.6	844	7	AAV60570
16	542	7.4	807	21	AAV54467
17	520.5	7.1	788	21	AAV54466
18	513	7.0	1837	21	AAV1726
19	506.5	6.9	744	9	AAV82975
20	501.5	6.9	1721	19	AAW48299
21	498.5	6.8	1721	21	AAV11727
22	497	6.8	182	12	AAV10872
23	496.5	6.8	2971	21	AAV41231
24	489.5	6.7	3118	22	AAV50363
25	489.5	6.7	2972	22	AAV50362
26	488.5	6.7	826	13	AAV26042
27	488	6.7	817	22	AAV16458
28	488	6.7	617	22	AAV04187
29	488	6.7	957	21	AAV59288
30	488	6.7	957	22	AAV24513
31	485	6.6	1127	22	AAV95541
32	476.5	6.5	652	9	AAV82974
33	467.5	6.4	511	22	AAV14883
34	467.5	6.4	511	22	AAV27312
35	467.5	6.4	511	22	AAV02607
36	456	6.3	2819	22	AAV35408
37	454.5	6.2	1012	20	AAV17406
38	450.5	6.2	378	12	AAV14160
39	448	6.1	1237	21	AAV81609
40	446.5	6.1	378	12	AAV14162
41	446.5	6.1	750	20	AAV05477
42	444.5	6.1	1532	21	AAV40945
43	444.5	6.1	3266	21	AAV42491
44	443.5	6.1	2685	22	AAV14533
45	443.5	6.1	2685	22	AAV26950

## ALIGNMENTS

RESULT 1	
AAV26049	standard; Protein: 1404 AA.
AAV26049	
02-FEB-1993	(first entry)
MSF precursor.	
Megakaryocyte colony stimulating factor; secretion signal; meg-CSF; stability; proteolytic cleavage; adhesion; alternative splicing.	
Synthetic.	
Location/Qualifiers	
1..26	/label= Exon_I
26..67	/label= Exon_II
67..107	/label= Exon_III
107..157	/label= Exon_IV
157..200	/label= Exon_V
200..1141	/label= Exon_VI
1411..1166	/label= Exon_VII
1166..1212	/label= Exon_VIII
1213..1266	/label= Exon_IX

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the Fa  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precu  
Cryptosporidium pa  
Portion of cryptos  
Protein encoded by  
Human ORFX ORF95  
Human SRAP. Homo  
P. yoe11 SSP2 ant  
Peptide #2869 enco  
Human MUC11 polype  
C900P predicted am  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
Human ORFX ORF709  
Human atrophin-1 r  
PRP 378. Trilicun  
Streptococcus pneu  
PRP encoded by clo  
C. albicans Rb1 p  
Human ORFX ORF225  
Human ORFX ORF225  
Peptide #967 enco  
Peptide #987 enco



QY 1223 FFRGSSIQOITYIKOEYOKCPGRRPAAINPVYGBMTQVRRRRERATGPGQHTTIRIQY 1282  
 DB 1261 ffrgssiqylykqepvqkpgrrpalnpyvgemtcvrrrrferrragspgtltiriy 1320  
 QY 1283 SPARLAVODKVLHNEKVSILMRGLPMVWVSATSLPIRKPDDYDYAFSKDDYINIDV 1342  
 DB 1321 sparlavqdyvlnmekvslmrtglpmvvsalslpilrpkpdydyaiskdylnldv 1380  
 QY 1343 PSRTARATITRSQGLTKRWYNCP 1366  
 DB 1381 psrtaratitrsqgltskwyncp 1404

RESULT 2  
 AAB60568  
 ID AAB60568 standard; Protein; 1404 AA.  
 AC AAB60568;  
 DT 27-APR-2001 (first entry)  
 DE Human megakaryocyte stimulating factor (MSF, CACP).  
 KW Human; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;  
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KW antiarthritic.  
 OS Homo sapiens.  
 PN WO200107068-A1.  
 PD 01-FEB-2001.  
 PE 21-JUL-2000; 2000MO-US20002.  
 PR 23-JUL-1999; 9905-0145328.  
 PR 19-JUL-2000; 200005-0145328.  
 PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 PI Warman ML;  
 DR MPI; 2001-182721/18.  
 PT New composition comprising the campodactylly-arthropathy-coxa  
 PT vara-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints  
 PS Example 1; Page 7; 34pp; English.  
 PS The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactylly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACP protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactylly-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACP protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACP protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U70316).  
 CC  
 CC Sequence 1404 AA;

Query Match 99.6%; Score 7265; DB 22; Length 1404;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1366; Conservative 0; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAKTLPIYLLLSLVYIQOVSQDLSACAGRGEGYSRDPATNCYINCHYMECCPDF 60  
 DB 1 mawtlpilyllllsvlrlqvsqdlsscagrcgysrdatcncynqhyameccpdf 60  
 QY 61 KRVCTA-----AEVHNPTSPSSSKAP 82  
 DB 61 krvctaelsckgicrfesfergrecedaqckkydkccpdyesfcaevhnpstpskkap 120  
 QY 83 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEEHSYSENOSSSSSSSSSTIIV 142  
 DB 121 ppsgasqtlstkrspkpnkktkkviiseeiteehsysengessssssstliw 180  
 QY 143 KIKSSKNSAANRELQKLLKYKDNKKNFTKKKPPKPPVVDAGSGLDNGDFKVTTPDTST 202  
 DB 181 kiksksnaaanrelqkllkykdnkknrtkkkppkppvvdagsgldngdfkvtlptdst 240  
 QY 203 TOHNVSTSPKITTAKEINRPSLPKSDPSKESLTVNKEETVETKETTNNQSTDG 262  
 DB 241 tqnhvstspkittakelpnrpslpnpsdstsketslvnketvettntnkqstldg 300  
 QY 263 KEKTTSAKEQSTIEKTSAKDLAPTSKVLAKPTPKAETTTGPAITTPKEPTTPKSPAS 322  
 DB 301 kektsakeqstiektakdlaptskvlakptpkaeittgpaaittpkepttpkspas 360  
 QY 323 TTPKEPTPTTIKSAPTTKKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEP 382  
 DB 361 ttpkeptpttiksaptttkkepapttkksaptttkkepapttkkepapttkkep 420  
 QY 383 APPTTKSAPTTKEPAPTTTKKRAPTTPKRAPPTPKRAPPTPKRAPPTPKRAPPTPK 442  
 DB 421 appttksapttkepapttkkrapttkrappttkrappttkrappttkrappttkp 480  
 QY 443 EPAPTAKKRAPTTPKRAPTTKKEPAPTTTKESPTTPKEPAPTTTKSAPTTKKEPAPT 502  
 DB 481 epaptakkrapttkrappttkkepapttkespttkkepapttkksaptttkkepapt 540  
 QY 503 TTKSAPTTKPESEPTTTKEPAPTTPKRAPTTPKRAPTTPKRAPTTPKRAPTTTKKP 562  
 DB 541 ttksapttkpespttkkepapttkkrapttkrappttkrappttkrappttkkp 600  
 QY 563 APAAPKRAPTTKEPAPTTPKRAPTTPKRAPTTPKRAPTTPKRAPTTPKRAPTTPEEPTT 622  
 DB 601 apaapkrapttkepapttkrappttkrappttkrappttkrappttkrapptt 660  
 QY 623 PEEAPPTPKRAAPNTPKRAPPTPKRAPPTPKRAPPTPKRAPPTPKRAPPTPKRAPTTKEP 682  
 DB 661 peepaptpkraapnptkrapttkrappttkrappttkrappttkrappttkrappttkp 720  
 QY 683 APPTPKRAPKEIAPTTTKPSTSDKRAPTTPKGAPTTPKRAPPTPKRAPPTPKG 742  
 DB 721 apttpkrapkeliaptttkpstsdkrapttkgapttkrappttkrappttkp 780  
 QY 743 TAPPTLKEPAPTTPKRAPKEIAPTTTKGPTSTSDKRAPTTPKRAPTTPKRAPPTPK 802  
 DB 781 tapptlkepapttkrapkeiaptttkgptstsdkrapttkrappttkrappttkp 840  
 QY 803 KPAPTPEPTPTTSEVSTPTTTKEPTTIKSPDESPPELSAEPPTKALNSPEEPVPT 862  
 DB 841 kpaptpeptpttsevsptttkepttikspdestpelsaepptkalnspeepvpt 900  
 QY 863 TKTPAAKPEMTTAAKDTTERDLKTTPETTTTAAPKMTKETAATTTKTESKTIATTTQY 922  
 DB 901 tktpaakpemttaaakdttterdlrttpetttaapkmktetattktesklatattqy 960  
 QY 923 TSTTTQDTTFKFTITLTKTTTLAKVYTTTKYTTTTEIMANKPEETAKPKDRATNSKATTPK 982  
 DB 961 tstttqdttfkftitltkttlapakvtttkkitttteimankpeetakpkdatnskattpk 1020



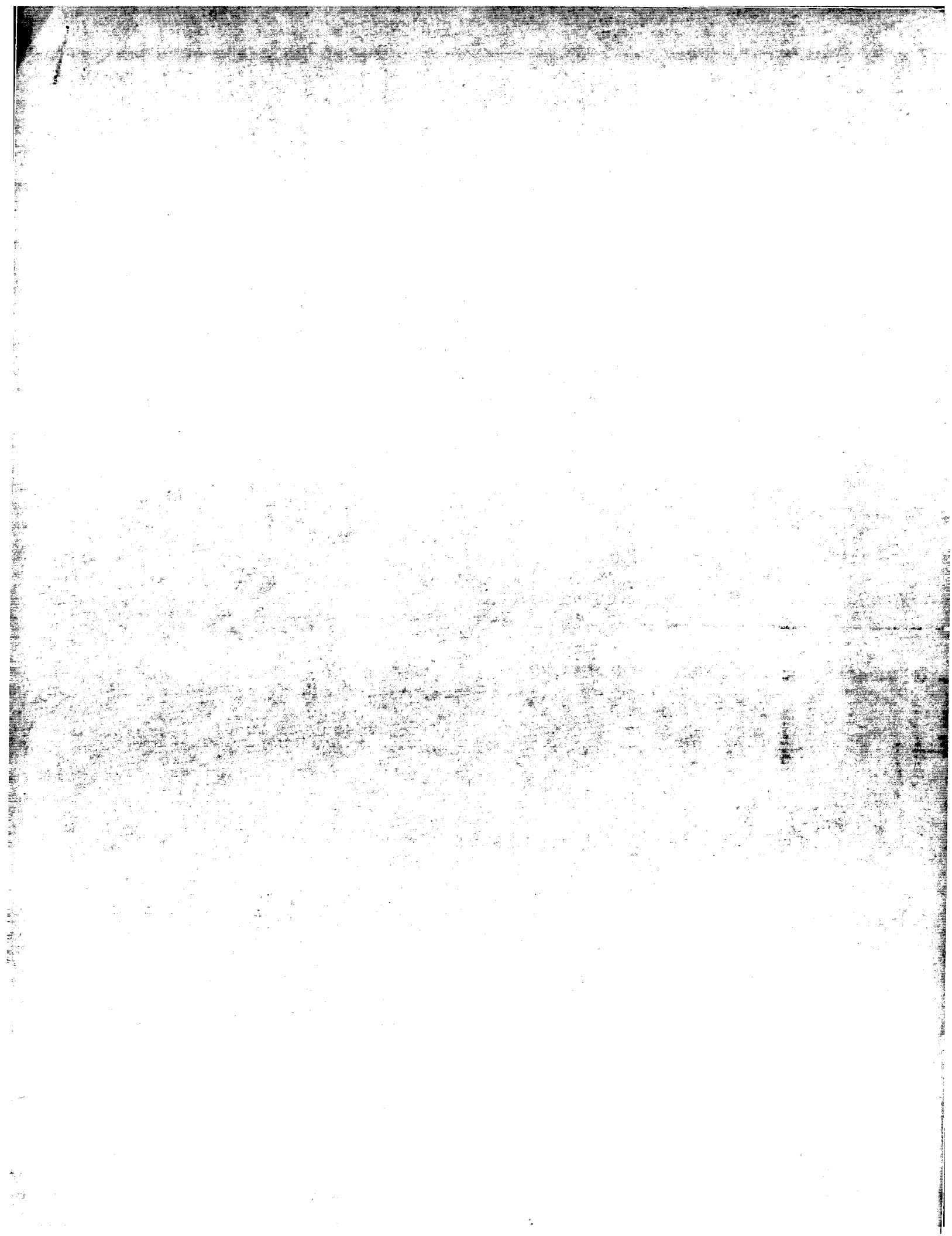


FT Region 1266..1331  
 FT /label= Exon\_X  
 FT Region 1331..1373  
 FT /label= Exon\_XII  
 FT Region 1373..1404  
 FT /label= Exon\_XII  
 PN WO9213075-A.  
 XX 06-AUG-1992.  
 PD 17-JAN-1992: 92MO-US00433.  
 XX  
 PR 18-JAN-1991: 91US-0643502.  
 PR 10-SEP-1991: 91US-0757022.  
 XX  
 PA (GEM ) GENETICS INST INC.  
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
 DR WPI: 1992-284660/34.  
 DR N-PSDB; AAQ27223.  
 XX  
 PT New human mega-karyocyte stimulating factors - for treating  
 PT immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.  
 PS  
 PS Claim 1, 2 and 3; Fig 1; 87pp; English.  
 CC The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine proceeding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC sequences derived from human megakaryocyte colony stimulating factor  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC primary transcript of this gene may be cleaved in different ways to  
 CC yield a family of mRNA's each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor, ie. one or more of these exons may contain  
 CC sequences which direct proteolytic cleavage, adhesion, organisation of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 XX  
 SO Sequence 1404 AA:

Query Match 97.6%; Score 7265; DB 13; Length 1404;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1366; Conservative 0; Mismatches 0; Indels 38; Gaps 1;

QY 1 MANKTLPYILLLLSVFIOQVSSQDLSGARGEGYSRDATNCMDYCNQHYMECCPDF 60  
 DB 1 maktlpiyllllsvfivqvsqdlscagrcggyrdatncndynqhymeccpdf 60  
 QY 61 KRVCIA-----AEVHNTPSPSSKAP 82  
 DB 61 krcvtaelsckgrcfesferecdagckkydkccpdyesfaehnpstpskcap 120  
 QY 83 PRGASQSTIKSTRSRSPKPKNKKTKKVISEETEEHSHVSENQSSSSSSSSSTIIV 142  
 DB 121 prgsaqlstikstrspkpkpkkkvieseeitshsvsenqssssssssstliw 180

QY 143 KIKSSKNSAANRELQKRLKVKDKNKKRKKPKPPVVDAGSGLDNGDFKVTTPDTST 202  
 DB 181 kiksksnsaanrelqkrlkvwkdnkkrtkkpkppvvdagsgjldngdfkvtptdst 240  
 QY 203 TQHNKYSTSPKITTAKPIVNPRLPNPSDSKTSLSLVNKEETVETKTTTAKKOSTDG 262  
 DB 241 tqhnkystspkittakpivnprrlpnpsdstsktslsvnketvetttkkqstldg 300  
 QY 263 KEKTSKAKETOSIEKTSANDLAPTSKVLAKPPTKAETTTGPAITTPKEPTTPPEPAS 322  
 DB 301 kektsaketosiektssandlaptskvlakpptaetttgpaittpkepttppepas 360  
 QY 323 TPPEKPTPTTIKSAPPTPKAPPTTKSAPPTPKAPPTTKKAPPTTKKAPPTTKK 382  
 DB 361 tppkepptpttiksapptpkappttksapptpkappttkkappttkkappttkk 420  
 QY 383 APPTTSAPPTPKAPPTPKKAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 442  
 DB 421 appttsapptpkapptpkkapptpkappttkkappttkkappttkkappttkk 480  
 QY 443 EPAP7APKRPAPPTPKAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAPPT 502  
 DB 481 epap7apkrpaptpkapptpkappttkkappttkkappttkkappttkkappttk 540  
 QY 503 TTKSAPPTPKESPPTTKKAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKK 562  
 DB 541 tksapptpkesppttkkapptpkappttkkappttkkappttkkappttkk 600  
 QY 563 AP7APKRPAPPTPKAPPTPKKLTPTTBKLAFTTPPKAPPTTBELAPTTBEEPTPT 622  
 DB 601 ap7apkrpaptpkapptpkkltpttbeklafttppkappttbelafttbeptpt 660  
 QY 623 PEEPA7PTTKAAPNPKPAPPTPKAPPTTKKAPPTTKKAPPTTKKAPPTTKKAP 682  
 DB 661 peepa7pttkaaapnkpapptpkappttkkappttkkappttkkappttkk 720  
 QY 683 APPTPKKAPKELAPPTTBEPSTSTSDKRAPPTTKAPPTTKKAPPTTKKAPPTPK 742  
 DB 721 aptpkkappkelappttbepststsdkrapttkappttkkappttkkapptpk 780  
 QY 743 TAPTTKEAPPTPKKAPKELAPTTTKGPTSTSGKAPPTTKKAPPTTKKAPPTTKK 802  
 DB 781 tapttkeapptpkkapkelaptttkgptstsgkappttkkappttkkappttkk 840  
 QY 803 KPA7PTPEPTPTSTSVSTPTTKKEPTTIHKSDEESTPELSAPPTKALENSKEEGVPT 862  
 DB 841 kpaptpeptptstsvstpttkkepttihksdestpelsapptkalenskeegvpt 900  
 QY 863 TKTPAATKPEMTTAKDRTTERDLRTTPTTAAPKMTKETATTEKTESKITATTVQY 922  
 DB 901 tktpaatkpemttakdrtterdlrttpttaapkmtketattekteskitattvy 960  
 QY 923 TSTTODTTPFKTTTLTKTTTAKVTTTKTTTITIMKPEETAPKXARATSKATTPK 982  
 DB 961 tstitodttpfkttltktttakvtttktttittimkpeetapkxaratskattpk 1020  
 QY 983 POKPTAPKPKSTKPKKTPMRVRKPKTTPPKMTSTMBELPTSRISAEAMLOTTRPN 1042  
 DB 1021 pokptapkpkstpkktpmrvrkpkttppkmtstmbelptsrisaeamlottrpn 1080  
 QY 1043 QTPNSKLVENPKSEDAGABGETPMMLLRPHVMEVTPPMDYDLRVPNOGIIINPMLS 1102  
 DB 1081 qtpnsklvenpkseadagaagetpmlrlphvmevtpmdydlrvpnogiiinpmis 1140  
 QY 1103 DETNIONGRVDELITLRNGILVAFRGHFWMLSPSPSPAPRRITVEWGISPDIPTVT 1162  
 DB 1141 detniongrvdelitlrngilvafrghfwmalspspsparritvewgispdptvt 1200  
 QY 1163 RCNCEGKTFEKKDSOYWRFTNDIKDAGYKPIFKGGGILGQIVAAIATAKYNMPESVY 1222  
 DB 1201 rcncegkttfkdsqywrftndikdagypkpfkfggilgqivaaiatakynmpevy 1260



SEA ID NO: 1, aab  
AC NO: AAR26049, A-Genaseg--1101

Mycobacterium tube  
Sequence g1/101742  
Mycobacterium tube  
Sequence of the fa  
Amino acid sequenc  
Amino acid sequenc  
Crypsosporidium pa  
Biothrestive precu  
Crypsosporidium pa  
Portion of Cryps  
Protein encoded by  
Human OREF ORF995  
Human OREF ORF995  
Human SRAP. Homo  
Human SRAP. Homo  
p. yeell1 SSP2 ant  
Peptide #2852 encd  
Peptide #2869 encd  
Human MUC11 polype  
G900P predicted am  
Human protein sequ  
Biothrestive precu  
Peptide #1317 encd  
Peptide #1349 encd  
Peptide #1285 encd  
Human O7GC27 gene  
Human atrophin-1 r  
PRP 378. Triticum  
Streptococcus pneu  
PRP encoded by clo  
C. albicans RBT1  
Human OREF ORF709  
Human OREF ORF2255  
Peptide #667 encd  
Peptide #867 encd  
Peptide #867 encd

655.691 million cell updates/sec

BLOSUM62

Total number of hits satisfying chosen parameters:	522463
--	--------

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	/SID8/gcgdata/genseq/genseqp/AAl980.DAT *
2:	/SID8/gcgdata/genseq/genseqp/AAl981.DAT *
3:	/SID8/gcgdata/genseq/genseqp/AAl982.DAT *
4:	/SID8/gcgdata/genseq/genseqp/AAl983.DAT *
5:	/SID8/gcgdata/genseq/genseqp/AAl984.DAT *
6:	/SID8/gcgdata/genseq/genseqp/AAl985.DAT *
7:	/SID8/gcgdata/genseq/genseqp/AAl986.DAT *
8:	/SID8/gcgdata/genseq/genseqp/AAl987.DAT *
9:	/SID8/gcgdata/genseq/genseqp/AAl988.DAT *
10:	/SID8/gcgdata/genseq/genseqp/AAl989.DAT *
11:	/SID8/gcgdata/genseq/genseqp/AAl990.DAT *
12:	/SID8/gcgdata/genseq/genseqp/AAl991.DAT *
13:	/SID8/gcgdata/genseq/genseqp/AAl992.DAT *
14:	/SID8/gcgdata/genseq/genseqp/AAl993.DAT *
15:	/SID8/gcgdata/genseq/genseqp/AAl994.DAT *
16:	/SID8/gcgdata/genseq/genseqp/AAl995.DAT *
17:	/SID8/gcgdata/genseq/genseqp/AAl996.DAT *
18:	/SID8/gcgdata/genseq/genseqp/AAl997.DAT *
19:	/SID8/gcgdata/genseq/genseqp/AAl998.DAT *
20:	/SID8/gcgdata/genseq/genseqp/AAl999.DAT *
21:	/SID8/gcgdata/genseq/genseqp/Aa2000.DAT *
22:	/SID8/gcgdata/genseq/genseqp/Aa2001.DAT *

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	DB	ID	Description
1	6470.5	98.5	1404	13	AA826049	MSF precursor. Sy
2	6470.5	98.5	1404	22	AA860568	Human megakaryocyt
3	6470.5	98.5	1404	22	AA829773	Human megakaryocyt
4	5894.5	89.7	1299	22	AA824332	Human EST encoded
5	3484	53.0	902	22	AA839778	Human MSF-related
6	1707.5	26.0	472	22	AA860569	Bovine MSF ortholo
7	1397.5	21.3	452	16	AA880041	Human megakaryocyt
8	1168	17.8	5179	22	AA824516	C899P predicted am
9	981	14.9	763	21	AA838992	Arabidopsis thalia
10	950	14.5	1664	19	AA843106	C. thermophilus O1
11	772.5	11.8	1325	22	AA803645	Peptide #3327 enco

12	715	10.9	753	18	AAW1852
13	625.5	9.5	441.2	21	AAV53666
14	612	9.3	572	18	AAW1855
15	553.5	8.4	844	7	AAV60570
16	542	8.3	807	21	AAV54467
17	520.5	7.9	788	21	AAV54466
18	508	7.7	1837	21	AAW17126
19	506.5	7.7	744	9	AAW82975
20	496.5	7.6	2971	21	AAW81231
21	489.5	7.5	2972	22	AAW80363
22	489.5	7.5	3118	22	AAW80362
23	489	7.4	1721	21	AAW11727
24	488.5	7.4	826	13	AAW60422
25	488	7.4	657	22	AAW64567
26	488	7.4	957	22	AAW04187
27	488	7.4	957	21	AAV52888
28	488	7.4	957	22	AAW24513
29	488	7.4	1721	19	AAW85299
30	485	7.4	1127	22	AAW85541
31	476.5	7.3	652	9	AAW82974
32	467.5	7.1	511	22	AAW14883
33	467.5	7.1	511	22	AAW27312
34	467.5	7.1	511	22	AAW026077
35	450.5	6.9	378	12	AAW41607
36	446.5	6.8	378	12	AAW4162
37	446.5	6.8	750	20	AAW54777
38	443	6.7	2870	21	AAV95559
39	443	6.7	3178	21	AAV95559
40	442.5	6.7	751	16	AAW80839
41	435.5	6.6	2819	22	AAW54083
42	431.5	6.6	2655	22	AAW24690
43	431.5	6.6	2655	22	AAW26950
44	431.5	6.6	2655	22	AAW02259
45	431.5	6.6	3266	21	AAW24291

## ALIGNMENTS

Myobacterium tub  
Sequence g1/101744  
Mycobacterium tub  
Sequence of the Fa  
Amino acid sequen  
Amino acid sequen  
Cryptosporidium p  
Blaohesive precu  
Human OREX ORF995  
Human SRCP. Hom  
Human SRCP. Hom  
Portion of Cryptoc  
P. yoelii SSP2 enc  
Peptide #892 enc  
Peptide #2869 enc  
Human MUC11 polyp  
G900P predicted a  
Cryptosporidium pr  
Human protein sequ  
Blaohesive precu  
Peptide #317 enc  
Peptide #1349 enc  
Peptide #1289 enc  
PRF 318. Trilium  
PRF encoded by C  
C. albicans Rb1T  
Caeirohhabitis el  
Caeirohhabitis el  
Japanese sea muss  
Human 07GC27 gene  
Peptide #667 enc  
Peptide #867 enc  
Peptide #941 enc  
Human ORFX ORF2255

RESULT	1
AA26049	
ID	AA26049 standard; Protein; 1404 AA.
XX	
AC	AA26049;
XX	
DT	02-FEB-1993 (first entry)
XX	
DE	MSF precursor.
XX	
KW	Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW	stability; proteolytic cleavage; adhesion; alternative splicing.
XX	
OS	Synthetic.
XX	
XX	
Key	Location/Qualifiers
FH	1..26
FT	/label= Exon_I
FT	26..67
FT	/label= Exon_II
FT	67..107
FT	/label= Exon_III
FT	107..157
FT	/label= Exon_IV
FT	157..200
FT	/label= Exon_V
FT	200..1141
FT	/label= Exon_VI
FT	1411..1166
FT	/label= Exon_VII
FT	1166..1212
FT	/label= Exon_VIII
FT	1213..1266
FT	/label= Exon_IX









QY 1086 EFKRGSIQOYIKOEPVCKGRRPALNYPVGEHTOVRRRRPERAIGSPQHTIRIQY 1145  
 DB 1261 ftkrsgslqylykqepvqkcprrlpmlpvygcmqvttrrrfetaispsqhtlrlqy 1320  
 QY 1146 SPARLKYODKGVLANEKKVSIWKGSLPNVNTSAISLPNKRPGDYDYAFSKDOYINIDY 1305  
 DB 1321 sparllyqdkgylnhekvslwrglpnvrtssalslpnlrkpdydyatskqynldv 1380  
 QY 1206 PSRTARAITRRSGOTLSKRWYNCP 1229  
 DB 1381 psrtaraitrrsgotlskwyncp 1404  
 RESULT 2  
 AABJ0568  
 ID AABJ0568 standard: Protein; 1404 AA.  
 AC AABJ0568;  
 XX  
 DT 27-APR-2001 (first entry)  
 DE Human megakaryocyte stimulating factor (MSF, CACF).  
 XX  
 KW Human: CACF protein; campodactyl-arthropathy-coxa vara-pericarditis;  
 KW MSF megakaryocyte stimulating factor; synovial lubricant;  
 KW chondrosarcoma 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KW arthralgia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MOJ0107068-01.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 21-FUL-2000; 2000MO-US20002.  
 XX  
 PR 23-JUL-1999; 99US-0145328.  
 PR 19-FUL-2000; 2000US-0145328.  
 XX  
 PA (UYPA-) UNIV CASE WESTERN RESERVE.  
 XX  
 PI Weiman ML;  
 XX  
 DR WPI; 2001-182721/18.  
 XX  
 PT New composition comprising the campodactyl-arthropathy-coxa  
 PT var-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints -  
 XX  
 PS Example 1; Page -: 34pp; English.  
 XX  
 CC The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the campodactyl-arthropathy-  
 CC coxa vara-pericarditis (CACF) protein, or portions of the CACF protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous  
 CC injection. The human CACF protein is identified in the invention as  
 CC being megakaryocyte stimulating factor (MSF). The gene encoding  
 CC CACF protein (MSF) is located on chromosome 1q25-31, and mutations in  
 CC this gene are responsible for the heritable disorder campodactyl-  
 CC arthropathy-coxa vara-pericarditis, in which patients have synovial  
 CC hyperplasia without evidence of inflammation. CACF protein (MSF)  
 CC acts as a synovium lubricant, and can be used to lubricate tissue and  
 CC joints in the treatment of osteoarthritis. The composition may be  
 CC applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
 CC loss of range of movement or joint damage). The present sequence  
 CC represents human megakaryocyte stimulating factor (MSF, CACF protein).  
 CC Note: This sequence is not given in its entirety in figure 4 of the  
 CC specification, although a GenBank accession number was given. This  
 CC sequence was therefore obtained from GenBank (U0316).  
 CC  
 SQ Sequence 1404 AA:

Query Match 98.5%; Score 6470.5; DB 22; Length 1404;  
 Best Local Similarity 87.5%; Pred. No. 0;  
 Matches 1229; Conservative 0; Mismatches 0; Indels 175; Gaps 1;  
 QY 1 MAWKTLPIYLLLLSYFVLIQOVS----- 24  
 DB 1 mawktlpiy111111svfivqvsqdlssacagcgysrdatcncdncqhyameccpdl 60  
 QY 25 ----- 24  
 DB 61 kryctaelsckgrcfesfergrecdcaqokkydkcodyesfcaevhnpstppsskka 120  
 QY 25 ----- 24  
 DB 121 ppegagqltksttkrpkppnkkkvlseeeitehsvsenqessssssssssslw 180  
 QY 25 -----VKDNKKNRKKKKPKPPVVDGSGGLNDGFKYTPPTST 65  
 DB 181 klksksnseennrelqkklvkdanknrtkkkppkppvdaagslndgfkvtcpdst 240  
 QY 66 TOHNKYSTSPKITTAKPINDPSLPPNSDTSKETSITVNRKETTVERKETTNNKQSTDC 125  
 DB 241 tqhkvstspkltakpindpslppnsdtsketsltvnrketvetttkqstsdg 300  
 QY 126 KEKTTSAKETQSIKTSKADLAPTSKYLAPTPKAETTTGPAITTPKEPTTPKREPA 185  
 DB 301 kektsaketqsietskadlaptskylvlaptkpaelttkgpaalttkpctpkpka 360  
 QY 186 TTPKEPTTPKSAPTTPKREPAITTPKSAPTTPKREPAITTPKREPAITTPKREPA 245  
 DB 361 ttpkepttpksapttpkrepa ittpksapttpkrepa ittpkrepa ittpkrepa ittpkrepa 420  
 QY 246 APTTTSAPTTPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPAITTPK 305  
 DB 421 aptttsapttkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 480  
 QY 306 EPAPTPAKKRPATTPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPA 365  
 DB 481 epaptpkapapttkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 540  
 QY 366 TTSAPTTPKREPSPTTPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPA 425  
 DB 541 tksapttkrepspttkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 600  
 QY 426 APTAPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPA 485  
 DB 601 aptapkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 660  
 QY 486 PEEPAITTPKRAAANPMPKREPAITTPKREPAITTPKREPAITTPKREPAITTPKREPA 545  
 DB 661 peepaittpkraaapnmpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 720  
 QY 546 APTTPKREPAKELAPTTTKPTSTSDKPAITTPKGAAPTPKREPAITTPKREPAITTPK 605  
 DB 721 apttpkrepa kelaptttkptstsdkpa ittpkgaaptpkrepa ittpkrepa ittpkrepa ittpkrepa 780  
 QY 606 TAPTLKEPAITTPKREPAKELAPTTTKGPTSTSDKPAITTPKREPAITTPKREPAITTPK 665  
 DB 781 taptlkepa ittpkrepa kelaptttkgptstsdkpa ittpkrepa ittpkrepa ittpkrepa ittpkrepa 840  
 QY 666 KPAITTPETPTTSEVSTTTTKEPTTIKSPDESIPESAEPTPALENSPEKEGYPT 725  
 DB 841 kpaittpetpttsevsitttkcptstsdspesaeptpalenspekegypt 900  
 QY 726 TKTPPAKRPMTTAKKKTTERDRTPEPTTAAKPKTKETATTEKTSKITAATTQV 785  
 DB 901 tktpaakrpmttacakktterdrtpeptttaa kpktketatektskitaattqv 960  
 QY 786 TSTTQDTPEFKITLTTTLTAPEVTTTKKTTTTEIMNKPEETAKKRDRAVNSKATTPK 845  
 DB 961 tsttqdtpefkittlittltapevtttkktttteimnkpeetakkrdra vnskattpk 1020



protein encoded by  
Myobacterium tube  
Sequence g1/201742  
MSF-K130. Synthetic  
Myobacterium tube  
Amino acid sequenc  
acid of the fa  
Cryptosporidium pa  
Blaedhesive precu  
Cryptosporidium pa  
Portion of Cryptos  
Human ORFX ORF995  
Human SRAP. Homo  
Human SRAP. Homo  
P. yoe11 S562 antib  
Peptide #2892 encod  
Peptide #2869 encod  
Human MUC11 Polype  
C900P predicted am  
Human protein sequ  
Blaedhesive precu  
Peptide #1317 encod  
Peptide #1349 encod  
Peptide #1289 encod  
Human\_07CG52 gene  
PRP 378. Trilicun  
Peptide #367 encod  
Peptide #867 encod  
Peptide #41 encod  
Human ORFX ORF225  
PRP encoded by chl  
C. albicans Rbt1-1  
Human atrophin-1



FT	Region	1266..1331
FT	/label= Exon_X	
FT	Region	1331..1373
FT	/label= Exon_XI	
FT	Region	1373..1404
FT	/label= Exon_XII	
XX		
XX	M09213075-A.	
XX		
PD	06-AUG-1992.	
XX		
PE	17-JAN-1992;	92WO-US00433.
XX		
PR	18-JAN-1991;	91US-0643502.
PR	10-SEP-1991;	91US-0757022.
PA	(GENY ) GENETICS INST INC.	
XX		
PI	Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;	
DR	MPf: 1992-284660/34.	
N	PSDB: AAO27223.	
PT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
PT	bacterial and viral infections, etc.	
XX		
XX	Claim 1, 2 and 3; Fig 1; 87pp; English.	
CC	The sequence given is a full length translation from the megakaryocyte	
CC	stimulating factor (MSF) precursor. The sequence covered by exons II,	
CC	III and IV encodes megakaryocyte stimulating factor (MSF). This	
CC	sequence is modified by the addition of an N-terminal sequence encoding	
CC	a secretory leader, an initiating methionine preceding exon II and a	
CC	terminating codon following exon IV. The cDNA sequence given contains	
CC	sequences derived from human megakaryocyte colony stimulating factor	
CC	(meg-CSF). Exon I contains the initiating methionine, and encodes a	
CC	classical mammalian protein secretion signal sequence. The sequence	
CC	encoding the original meg-CSF includes exons II-IV and is thought to	
CC	terminate in the region between amino acid residues 134 - 147. The	
CC	primary transcript of this gene may be cleaved in different ways to	
CC	yield a family of mRNAs each encoding a different MSF protein. Exons	
CC	V and VI are thought to be related to the activity of the factor and	
CC	are also implicated in the stability, folding and processing of the	
CC	molecule. These exons are also thought to play a role in the observed	
CC	synergy of MSF with other cytokines. Exons V - XII are believed to be	
CC	implicated in the processing or folding of the appropriate structure of	
CC	the resulting factor, ie. one or more of these exons may contain	
CC	sequences which direct proteolytic cleavage, adhesion, organisation of	
CC	the cellular matrix or extracellular matrix processing. Both naturally	
CC	occurring and non-naturally occurring MSF's may be characterised by	
CC	various combinations of alternatively spliced exons from this sequence,	
CC	with the exons spliced together in differing orders to form different	
CC	members of the MSF family.	
XX		
SQ	Sequence 1404 AA:	
Query Match	99.6%; Score 7391.5; DB 13; Length 1404;	
Best Local Similarity	96.9%; Pred. No. 0;	
Matches 1361: Conservative	0; Mismatches 0; Indels 43; Gaps	
OY	1AAAKTLPLTLLLSVTVQGVSDQLSCAGRCGEGRYDATNCNDVCQHMECCPDF 60	
DJ	1MAWKLPILYIILLISVFIVIGVSQGLSCACAGCGEGRYRDATCNDVCQHMECCPF 60	
OY	KRVCAAEISCRCRCESEFERGCCDDACKKRYDKCCCPYEFCFAVHNTSPSSKKAP 120	
DJ	KRVCAAEISCRCRCESEFERGCCDDACKKRYDKCCCPYEFCFAVHNTSPSSKKAP 120	
OY	PPSGASQRIRKSTTKTSKPKPNKKTKTKVIESEITE----- 156	
DJ	PPSGASQRIRKSTTKTSKPKPNKKTKTKVIESEITE----- 156	
OY	121 PPSGASQRIRKSTTKTSKPKPNKKTKTKVIESEITE----- 156	
DJ	121 PPSGASQRIRKSTTKTSKPKPNKKTKTKVIESEITE----- 156	

QY	157	-	-----VKDKKRRRTKKRPKRPVYDEAGSGJDNDDFVTPRPNST	197
Db	181	k1ksksnaanezqlckllyvdkhknrrtkkcpbpvndeagsqlandgflvccpdcst	240	
QY	198	TOHNKYSTSEKITTAAPINPRPSLPNSDSTKETSJLVNKEETVTEKETTTNNKQSTDG	257	
Db	241	tqhnkvastpklctakpndrpslpnsdstsketsjlvnketvtekettnknqtsldg	300	
QY	258	KERTSAKKEQOSLEKTSADDLAPTskvLAKPPPKAETTTKGALTTPKERTYTTKREPAS	317	
Db	301	kektsaeketqsketsakdlaptsakvLakppkeettkgsalctpkpctcpkpas	360	
QY	318	TTTTPEEPTTTKAPPTPKKEBAPTTTKSAPTPKKEBAPTTEKBPATPKKEBAPTTTKEP	377	
Db	361	ctpkpctcpkksapctpkpkeapcttkcsapctpkpkeapcttkkeapcttkpkeapcttkp	420	
QY	378	APTTSKAPTTPKEBAPTTPPKKAPPTPKKEBAPTTPPKKEBPTTPPKKEBAPTTPKEBAPTTPK	437	
Db	421	apcttkkapctpkpkeapcttpkpkapcttkpkeapcttkpcttkpkeapcttkpkeapcttkp	480	
QY	438	EPAPATKAKAPPTPKKEBAPTTPPKKEBAPTTPPKESPTTPKEBAPTTPKSAPTTTKEBAPT	497	
Db	481	epapctpkpkapctpkpkeapcttkpkeapcttkespttkpkeapcttkksapcttkkeapct	540	
QY	498	TTKSAPTTPKEEPTTTKKEBAPTTPKEBAPTTPPKKAPPTPKKEBAPTTPPKKEBAPTTTKKP	557	
Db	541	ctksapctpkpkeapcttkkeapcttkpkeapcttkpkpkapcttkpkeapcttkpkeapcttkp	600	
QY	558	APTAPKEBAPTTPTKRETAPPTTPPKKLAPTTPEBKLAAPTPPEKAPPTPEBELAPTPPEEPTPT	617	
Db	601	apctpkpkeapcttkpkeapcttkpklltpctpkpkeapcttkpkeapcttkpeelapcttkpeepct	660	
QY	618	PEEPAPTTPKAAAPNTPKKEBAPTTPPKKEBAPTTPPKKEBAPTTPPKETAPPTPKGAPTTLKEP	677	
Db	661	peepapcttkpkaapntpkpkeapcttkpkeapcttkpkeapcttkpkeapcttkpctkgacttkkcp	720	
QY	678	APTTPPKKAPKELAPPTTKKEPTSTSDKAPPTPKGATAPTKKEBAPTTPKREBAPTTPKKG	737	
Db	721	apcttpkpkpakeapcttkkeapcttkesctsdkpacpttkpctpkpkeapcttkpkeapcttkp	780	
QY	738	TAPTTLKEBAPTTPPKKAPKELAPPTTKGPTSTSDKAPPTPKKETAPPTPEEBAPTTPK	797	
Db	781	tapcttkkeapcttkpkeapcttkpkpcttkkpsctskspacttkpkeapcttkpkeapcttkp	840	
QY	798	KPAPTTPETPTTSEVSTPTTTKEBPTTIHKSPTDESPTPELSAEPPTKALENSPKBGPVT	857	
Db	841	kpacpttkpcttksevspttkcttkpcttkhksptdesptksaepctkalsenpkpkegpyct	900	
QY	858	TKTPPAAKKPEMTTATKDKTTERDLRTTPEPTTTAPAPPMKETEATTTBEKTSKITATTOY	917	
Db	901	tktpaakkepmittakdkctterdlrttpepttktaapmketatekctsktskittactty	960	
QY	918	TSTTQOTPEPFKTTTTKTTTTLAKVYTTTKKTIITTEIMNKPEETAKPKORANSNATPK	977	
Db	961	tstttqcttpcfklttltcttlapkvttcttklttcttsimkpeetakpkoranasnactpk	1020	
QY	978	PQKPTKAPKPTSTTKRKPTMPRVKRPKRTTTPPKMTSTMPBELNPTSRILAEAMLOTTTRN	1031	
Db	1021	pqkptkappkptsttkpkptmpvrvtkpcttkpkmstmpelnptsrilaeamlqctttrpn	1084	
QY	1038	QTPMSKVLVENPKSESDAGABEGSTPHMLRPHVMEVTPDMODYLRYVNOGIIINPKLS	1091	
Db	1081	qtpmskvlvenpkseadagabegstphmlrphvmevtpdmodylrvpnoqiiinpmis	1144	
QY	1098	DETNIQCKRGVDDGTLTLRNCTLVAFEGHFWMLSPSPSPSPARITTEWGTISPDYVTT	1157	
Db	1141	detniqckrgvddgltltnqcltvaftgghfwmlspspsparrittewgtspsdyvt	1200	
QY	1158	RCNCEGTFFFKSOYWRFNNDKDGYPKPLFKFGGGLTGOIVALSATKAKKMPESY	1211	
Db	1201	rcncegtfffksgywrfnndkdgypkplfkfgggltgolvalsaatkakmpesy	1264	



QY 1218 FFRGGSIQOYIYKQEPVOKCPGRPALNYPVYGEOTVRRRFRERAIQSPQTHIRIQY 1277  
 DB 1261 ffrgsgsiqylykgepvokcpgrpalnypvygeotyrrrrfreraiqspqthiriry 1320  
 QY 1278 SPARLAQOKGVILHNEVKVSIIMRGJLPNVVTSISLPIIRKRGDGDYAFPSQDYNNIDV 1337  
 DB 1321 sparlayqdkgyvlhnevkvsiimrgjlpnvvtalsipnirkpdydyafafskdyynidv 1380  
 QY 1338 PSTARAITRRSGQTLRSKVMNCP 1361  
 DB 1381 pstaraitrrsgqtlrskvynncp 1404

RESULT 2

AA0568  
 ID AAB60568 standard; Protein: 1404 AA.

AC AAB60568;

DT 27-APR-2001 (first entry)

DE Human megakaryocyte stimulating factor (MSF, CACP).

KW Human: CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;  
 MSF; megakaryocyte stimulating factor; synovial lubricant;

KM Chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 arthritis; etc.

KX Homo sapiens.

OS Homo sapiens.

PN W0200107068-AA.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000NO-US20002.

PR 23-JUL-1999; 9806-0145328.

PR 19-JUL-2000; 2000US-0145328.

XX (DYCA-) UNIV CASE WESTERN RESERVE.

PA (DYCA-) UNIV CASE WESTERN RESERVE.

PI (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

PT (DYCA-) UNIV CASE WESTERN RESERVE.

Query Match 99.6%; Score 7291.5; DB 22; Length 1404;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMTLPFYLLLLSYFVLIQOVSODLSSCAGRGEGSRATONCYNCHYMECCDF 60  
 DB 1 MAMTLPFYLLLLSYFVLIQVSSQDLSSCAGRGEGSRATONCYNCHYMECCDF 60  
 QY 61 KVCYTAELSCGRGCFESFERGECDCAOCKKYDCPDYEFCAEYHNPSPSSKAP 120  
 DB 61 KVCYTAELSCGRGCFESFERGECDCAOCKKYDCPDYEFCAEYHNPSPSSKAP 120  
 QY 121 PPSGASQTIKSTTRRSKRPKPKKTKYVISEETFE----- 156  
 DB 121 PPSGASQTIKSTTRRSKRPKPKKTKYVISEETFE----- 156  
 QY 157 -----VADNKNRTRKKKPKPKPVYDEAGSLGNDGKVTPTOST 197  
 DB 181 KIKESKNSAANRELQKLLKVDKNNKTKKPKPVYDEAGSLGNDGKVTPTOST 240  
 QY 198 TQHNKVESTPKITAKPDPNPSPSLPNSDTSKETSILTVNKEVETRETTTNRQSTDG 257  
 DB 241 TQHNKVESTPKITAKPDPNPSPSLPNSDTSKETSILTVNKEVETRETTTNRQSTDG 300  
 QY 258 KERTSAKENSSTKTSKADLAPSKYLAPTPKAETTTGPAITTRKEPTTPPKEPAS 317  
 DB 301 KERTSAKENSSTKTSKADLAPSKYLAPTPKAETTTGPAITTRKEPTTPPKEPAS 360  
 QY 318 TTPKEPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEP 377  
 DB 361 TTPKEPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEP 420  
 QY 378 APPTTKSAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPK 437  
 DB 421 APPTTKSAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPK 480  
 QY 438 EPAPTAKKRPAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTP 497  
 DB 481 EPAPTAKKRPAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTP 540  
 QY 498 TTKSAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEP 557  
 DB 541 TTKSAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEP 600  
 QY 558 APPTAKKRPAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTP 617  
 DB 601 APPTAKKRPAPTPPKEPASPTTPPKEPASPTTPPKEPASPTTPPKEPASPTTP 660  
 QY 618 PEEPAPTPPKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEPAPTPPKEP 677  
 DB 661 PEEPAPTPPKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEPAPTPPKEP 720  
 QY 678 APPTPKKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEPAPTPPKEP 737  
 DB 721 APPTPKKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEPAPTPPKEP 780  
 QY 738 TAPPTLKEPAPTPPKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEP 797  
 DB 781 TAPPTLKEPAPTPPKRAAPNPMPKEPAPTPPKEPAPTPPKEPAPTPPKEP 840  
 QY 798 KPAPTTPPPPTSEVSTPTTKKEPTTIKSPDESPELSAPTPPKALENSKREPVPT 857  
 DB 841 KPAPTTPPPPTSEVSTPTTKKEPTTIKSPDESPELSAPTPPKALENSKREPVPT 900  
 QY 858 TKPAPATKPPMTTAAOKTTERDLRTPEPTTAAAPKREVAATTTKTESKITTATTOY 917  
 DB 901 TKPAPATKPPMTTAAOKTTERDLRTPEPTTAAAPKREVAATTTKTESKITTATTOY 960  
 QY 918 TSTTODTPPEFTTTLTKTTTLAPKVTTTTKTTTITIMKPEPTAKPKDRAVNSKATTPK 977  
 DB 961 TSTTODTPPEFTTTLTKTTTLAPKVTTTTKTTTITIMKPEPTAKPKDRAVNSKATTPK 1020





RESULT	1	
AA826049	standard; Protein; 1404 AA	
ID	AA826049	standard; Protein; 1404 AA
XX		
AC	AA826049;	
XX		
DT	02-FEB-1993	(first entry)
XX		
DE	MSF precursor.	
XX		
XX		
KM	Megakaryocyte colony stimulating fa	
KM	stability; proteolytic cleavage; ad	
XX		
OS	Synthetic.	
XX		
FH	Key	
FT	Region	Location/Qualifiers
FT	Region	1..26
FT	Region	/label= Exon_I
FT	Region	26..67
FT	Region	/label= Exon_II
FT	Region	67..107
FT	Region	/label= Exon_III
FT	Region	107..157
FT	Region	/label= Exon_IV
FT	Region	157..200
FT	Region	/label= Exon_V
FT	Region	200..1141
FT	Region	/label= Exon_VI
FT	Region	1411..1166
FT	Region	/label= Exon_VII
FT	Region	1166..1212
FT	Region	/label= Exon_VIII
FT	Region	1213..1266
FT	Region	/label= Exon_IX



FT Region 1266..1331  
 FT /label- Exon\_X  
 FT 1331..1373  
 FT /label- Exon\_XI  
 FT 1373..1404  
 FT /label- Exon\_XII  
 PN W09213075-A.  
 XX  
 PD 06-AUG-1992.  
 XX  
 PF 17-JAN-1992: 92W0-US00433.  
 XX  
 PR 18-JAN-1991: 91US-0643502.  
 PR 10-SEP-1991: 91US-0757022.  
 XX  
 PA (GENY) GENETICS INST INC.  
 XX  
 PI Clark SC, Gesner TG, Hewlck RM, Jacobs K, Turner K;  
 XX  
 DR W01: 1992-284660/34.  
 DR N:PSDB:AA027223.  
 XX  
 PT New human mega-karyocyte stimulating factors - for treating  
 PT immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.  
 XX  
 PS Claim 1, 2 and 3; Fig 1: 87pp: English.  
 XX

CC The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine preceding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC sequences derived from human megakaryocyte colony stimulating factor  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC primary transcript of this gene may be cleaved in different ways to  
 CC yield a family of mRNA's each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor, ie. one or more of these exons may constrain  
 CC sequences which direct proteolytic cleavage, adhesion, organisation of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 XX

Sequence 1404 AA:

Query Match 99.2%; Score 7007.5; DB 13; Length 1404;  
 Best Local Similarity 93.4%; Pred. No. 0;  
 Matches 131; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAMTTLPTLLLSVYIOVSSODLSSCAGRCGEGSRDATCNCDYNCQHMECCPDF 60  
 DB 1 mawtlptllllsvfivqvsqdlsscagrcgsgysratcdncdynqchmecpdf 60  
 QY KRVTAETSCGRCGEGSEREGEDCAOCCKKYDKCCPDYSSFCAE----- 106  
 DB 61 krvtaetscgrcgcsfesteriedcdagckkkydkccpdyesfcaevhnptspsskkap 120  
 QY 107 ----- 106  
 DB 121 ppsgasqlketkrspkpkpkkttkrvleaeelteehsvsengessssssssstliw 180

QY 107 -----VKDNKKNRKKRKKRPKPVVDAGSGLDNGDKRYTTPNST 147  
 DB 181 kksksksaanrelqkklxvdknkrkkkpkpvpvdaagsglndgdkkvtpdstc 240  
 QY 148 TOHNVSTSPKITTAAPINRPSPSLPNSDTSKEPSTLVNKEETVEETTKTNKQTSIDG 207  
 DB 241 tqhkvstspkittaklmpnpslppnsdtskesslvnketvettcttnqtsidg 300  
 QY 208 KEKTSKAKETOSIEKTSKADLAPTSKVLAKPRAETTTKGALVTPKEPPTPKEPAS 267  
 DB 301 kektsaketqslektsakdlaptskvlakpbraetttkgalvtpkepptpkepas 360  
 QY 268 TTPKEPPTTIKSAPTTPKEAPPTTTSAPTPKEAPPTTKEAPPTTKAPPTTKEP 327  
 DB 361 ttpkeppttiksaptpkeapptttsaptpkeappttkeappttkappttkep 420  
 QY 328 APPTTSAPPTPKAPPTTPKAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTPK 387  
 DB 421 appttsapptpkappttpkeappttkeappttkeappttkeappttkapptpk 480  
 QY 388 EPAPTAAPKAPATTPKAPATTPKAPATTPKESPTTPKESAPATTPKSAPTTKAPAPT 447  
 DB 481 epaptaapkapattpkeappttpkeappttpkesspttpkessapattpkapapt 540  
 QY 448 TTKSAPTPKESPTTTKEAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKKP 507  
 DB 541 tksaptpkesptttkeapattpkapattpkapattpkapattpkapattpkap 600  
 QY 508 APATPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPK 567  
 DB 601 apatpkapattpkapattpkapattpkapattpkapattpkapattpkapattpk 660  
 QY 568 PEEPAPTPPAAANPPTPKESAPPTTPKESAPPTTPKESAPPTTPKESAPPTTPKES 627  
 DB 661 peepattppaaanpptpkessappttpkessappttpkessappttpkessap 720  
 QY 628 APPTPKAPAKELAPTTTKEPSTSTSKAPATTPKGAATTPKAPATTPKESAPPTPKG 687  
 DB 721 aptpkapakelaptttkepststskapattpkgaattpkapattpkessappt 780  
 QY 688 TAPPTLEAPATTPPKAPAPKELAPTTTKEPSTSTSDKAPATTPKETAAPTPKESAPPTPK 747  
 DB 781 tapptleapattppkapapkelaptttkepststsdkapattpketaptpkeapattpk 840  
 QY 748 KPATTPETPTPTSEVSTPTTKEPTTKSPDESPELSAEPKALENSPEKEPVT 807  
 DB 841 kpattpetptptsevstpttkepttkspdespelasaepkalenspekapt 900  
 QY 808 TKTPPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTEKTSKIRATTTQV 867  
 DB 901 tktpaatkpemttakdkterdlrttpeettaapkmktetatttekteskiratctq 960  
 QY 868 TSTTQOTTPFKITTLKTTTLAPKVYTTTKTITTTETLMMKPEETAKKDRATNSKATTPK 927  
 DB 961 tsttqottpfkittllktttlapkvtttkitttletlmmkpeetakkdraztnakatpk 1020  
 QY 928 POKPTKAPKPTSTTKRPTPRVAKKPTTPPRKSTSTMPDELNTSLAEMLOTTTRPN 987  
 DB 1021 pokptkapkptsttkrptprvakprrptprkststmpdelntslaeamlotttrpn 1080  
 QY 988 QTPNSKLVEYNPKSEDAAGAGETPHMLLRPHVMPETVTDMDLPRVNPNOGIINPMLS 1047  
 DB 1081 qtpnslveynpksedagagetphmlrrphvmpetvtdmdlprvnpngiinpmls 1140  
 QY 1048 DENNICNGKRVYDGLITLRNGTLVAFRGHYTWMLSPPSPSPAPARITVWGIIPSIDVTFT 1107  
 DB 1141 dennicngkryvdyglitlrngtlvafrghytwmllspspspaparitvewgiipsidvtft 1200  
 QY 1108 RCNCEGKTFEPFKDSQYVRFNDIKDAGYPRKIFKRGEGGLGOIYAAALSTKRYKNWPSYV 1167  
 DB 1201 rcncegkttffkdsqyvrftndikdagyprrkifkrgggllgoiyaaalstakryknwpsyv 1260



Query Match	99.2%	Score 7007.5	DB 22	Length 1404
Best Local Similarity	93.4%	Pred. No. 0		
Matches 1311	Conservative 0	Mismatches 0	Indels 93	Gaps 1
QY 1	MAKMTLPIYLALLSVFIQQVSSODLSSCGRGSGESYRDACNCDYKQAHMECCPDF	60		
Db 1	MAKMTLPIYLALLSVFIQQVSSODLSSCGRGSGESYRDACNCDYKQAHMECCPDF	60		
QY 61	KRYTAEISCKGRCPESEREREDCDACGKRYDKCCPDYESFCAL	106		
Db 61	KRYTAEISCKGRCPESEREREDCDACGKRYDKCCPDYESFCAL	106		
QY 107	PPSGASGLKSTCRSPKPKKKTKKVLSSSEELTEEHAVSENGEASSSSSSSSSSLIW	180		
Db 107	PPSGASGLKSTCRSPKPKKKTKKVLSSSEELTEEHAVSENGEASSSSSSSSSSLIW	180		
QY 181	KIKSSKNSAAATELQKKLKYVDKKNRTKKKPKPKPVVDGSGDNGDFVTPPDT	147		
Db 181	KIKSSKNSAAATELQKKLKYVDKKNRTKKKPKPKPVVDGSGDNGDFVTPPDT	147		
QY 148	TOHNVKSTSPKIDAPINRPSLPSNDSOTSKTSLTVNKEVTEVETKETTNNKOTSDG	240		
Db 148	TOHNVKSTSPKIDAPINRPSLPSNDSOTSKTSLTVNKEVTEVETKETTNNKOTSDG	240		
QY 241	TQHNKSTSPKIDAPINRPSLPSNDSOTSKTSLTVNKEVTEVETKETTNNKOTSDG	207		
Db 241	TQHNKSTSPKIDAPINRPSLPSNDSOTSKTSLTVNKEVTEVETKETTNNKOTSDG	207		
QY 208	KEKTSKKEFQSIETKYNADLAPTSKVLKPKPKAETTTKGALTPKEPMTPTTKERAS	267		
Db 301	KEKTSKKEFQSIETKYNADLAPTSKVLKPKPKAETTTKGALTPKEPMTPTTKERAS	267		
QY 268	TPPEPPTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	327		
Db 361	TPPEPPTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	327		
QY 328	APPTTSAPPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	387		
Db 421	APPTTSAPPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	387		
QY 388	EPAPTPPKKAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAP	447		
Db 481	EPAPTPPKKAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAP	447		
QY 448	TKKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKER	507		
Db 541	TKKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKER	507		
QY 508	APCAPKEPAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAP	567		
Db 601	APCAPKEPAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAP	567		
QY 568	PEEPAPTPPKAAAPTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	627		
Db 661	PEEPAPTPPKAAAPTKERAPATTTKSAPTTKERAPATTTKSAPTTKERAPATTTKSA	627		
QY 628	APPTTKKAPKELAPATTTKEPTSTSDKAPATTTKGAATPTTKERAPATTTKGA	687		
Db 721	APPTTKKAPKELAPATTTKEPTSTSDKAPATTTKGAATPTTKERAPATTTKGA	687		
QY 688	TAPPTTKKAPKELAPATTTKEPTSTSDKAPATTTKGAATPTTKERAPATTTKGA	747		
Db 781	TAPPTTKKAPKELAPATTTKEPTSTSDKAPATTTKGAATPTTKERAPATTTKGA	747		
QY 748	KPAPTPEPTPEPTSEVSTPTTKKEPTTHKSDPDSSTSELNAPTKAKNSPREPVP	807		
Db 841	KPAPTPEPTPEPTSEVSTPTTKKEPTTHKSDPDSSTSELNAPTKAKNSPREPVP	807		
QY 808	TKTPAAKPKEMTTAKDOKTERDLRTPEPTTAAPKMTKEPATTTKSTKTTATTTQV	867		
Db 901	TKTPAAKPKEMTTAKDOKTERDLRTPEPTTAAPKMTKEPATTTKSTKTTATTTQV	867		
QY 868	TSSTODTPPKTTTKKTTTLAPKTTTKKTTTTELMNPEETAKKDAATNSKATYTK	927		
Db 927	TSSTODTPPKTTTKKTTTLAPKTTTKKTTTTELMNPEETAKKDAATNSKATYTK	927		



1

OM protein - protein search, using sw model

Title:	AA2
Perfect score:	7276
Sequence:	1 MAMKTLPIYLLLSVFVIQ.....ARAITTRSGQTISKVWYNCP 1363

**BLOSUM62**

Total number of hits satisfying chosen parameters: 522463

Listing first 45 summaries

1:	/SIDS8/gcgdata/genseq/genseq/A11980.DAT *
2:	/SIDS8/gcgdata/genseq/genseq/A11981.DAT *
3:	/SIDS8/gcgdata/genseq/genseq/A11982.DAT *
4:	/SIDS8/gcgdata/genseq/genseq/A11983.DAT *
5:	/SIDS8/gcgdata/genseq/genseq/A11984.DAT *
6:	/SIDS8/gcgdata/genseq/genseq/A11985.DAT *
7:	/SIDS8/gcgdata/genseq/genseq/A11986.DAT *
8:	/SIDS8/gcgdata/genseq/genseq/A11987.DAT *
9:	/SIDS8/gcgdata/genseq/genseq/A11988.DAT *
10:	/SIDS8/gcgdata/genseq/genseq/A11989.DAT *
11:	/SIDS8/gcgdata/genseq/genseq/A11990.DAT *
12:	/SIDS8/gcgdata/genseq/genseq/A11991.DAT *
13:	/SIDS8/gcgdata/genseq/genseq/A11992.DAT *
14:	/SIDS8/gcgdata/genseq/genseq/A11993.DAT *
15:	/SIDS8/gcgdata/genseq/genseq/A11994.DAT *
16:	/SIDS8/gcgdata/genseq/genseq/A11995.DAT *
17:	/SIDS8/gcgdata/genseq/genseq/A11996.DAT *
18:	/SIDS8/gcgdata/genseq/genseq/A11997.DAT *
19:	/SIDS8/gcgdata/genseq/genseq/A11998.DAT *
20:	/SIDS8/gcgdata/genseq/genseq/A11999.DAT *
21:	/SIDS8/gcgdata/genseq/genseq/A12000.DAT *
22:	/SIDS8/gcgdata/genseq/genseq/A12001.DAT *

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	Score	Query No.	Match	Length	DB	ID	Description
1	7245.5	99.6	1004	13	AA826049		MSF precursor. Sy
2	7245.5	99.6	1404	22	AA860568		Human megakaryocyt
3	7245.5	99.6	1404	22	AA829773		Human megakaryocyt
4	6669.5	91.7	1299	22	AA824322		Human EST encoded
5	3484	47.9	902	22	AA828748		Human MSF-derived
6	1916.5	26.3	452	16	AA880041		Human megakaryocyt
7	1707.5	23.5	472	22	AA860569		Bovine MSF ortholog
8	1175	16.1	5179	22	AA828516		C8939 predicted am
9	981	13.5	763	21	AA838942		Arabidopsis thaliana
10	950	13.1	1664	19	AAW43106		C. thermocellum O1
11	805	11.1	1325	22	AAW03645		peptide C12327 enco

File copy

12	715	9.8	763	18	AAW31852
13	645.5	8.9	444.2	21	AAV53666
14	612	8.4	572	18	AAV51855
15	555.5	7.6	884	7	AAW60570
16	542	7.4	807	21	AAV54467
17	520.5	7.2	768	21	AAV54466
18	503	7.1	1837	21	AAV11726
19	517	7.0	182	12	AAW10872
20	506.5	7.0	744	9	AAW82975
21	498.5	6.9	1721	19	AAW17737
22	468	6.8	1271	19	AAW48299
23	496.5	6.8	2971	21	AAW80331
24	489.5	6.7	2972	22	AAW50123
25	489.5	6.7	3118	22	AAW50362
26	488.5	6.7	826	13	AAW56042
27	488	6.7	617	22	AAW16458
28	488	6.7	617	22	AAW04187
29	488	6.7	957	21	AAV59288
30	488	6.7	957	22	AAW24513
31	485	6.7	1137	22	AAW95541
32	476.5	6.5	652	9	AAW2974
33	467.5	6.4	511	22	AAW14863
34	467.5	6.4	511	22	AAW27312
35	467.5	6.4	511	22	AAW02607
36	454.5	6.2	1012	20	AAV17406
37	454	6.2	2819	22	AAW35408
38	450.5	6.2	378	12	AAW14160
39	448	6.2	1237	21	AAW41609
40	447.5	6.2	326	21	AAW42491
41	446.5	6.1	378	12	AAW14162
42	446.5	6.1	750	20	AAV05474
43	446.5	6.1	265	22	AAW14533
44	446.5	6.1	265	22	AAW26950
45	446.5	6.1	265	22	AAW02259

## RESULT

ID AAR26049 standard; Protein; 1404 AA.

AC AAR260497

DT 02-FEB-1993 (first entry)

DE MSF precursor.

KW Megakaryocyte colony stimulating factor: meg-CSF.

stability; proteolytic cleavage; adhesion; alternative splicing.

Synthetic.

FH	Key	Location/Qualifiers
FT	Region	1..26

## Estimation

Region  
ET

Region

ET

Region

Reaction

ET

FT Region

33

Region

Region

ET

Region

### I.3





FT Region 1266..1331  
 FT /label= Exon\_X  
 FT 1331..1373  
 FT /label= Exon\_XI  
 FT 1373..1404  
 FT /label= Exon\_XII  
 PN W09213075-A.  
 PD 06-AUG-1992.  
 XX  
 PF 17-JAN-1992: 92WO-US00433.  
 XX  
 PR 18-JAN-1991: 91US-0643502.  
 PR 10-SEP-1991: 91US-0757022.  
 XX  
 PA (GENM ) GENETICS INST INC.  
 XX  
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
 XX  
 DR WPI: 1992-284660/34.  
 DR N-PSDB: AAQ27223.  
 XX  
 PT New human mega-karyocyte stimulating factors - for treating  
 PT immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.  
 XX  
 PS Claim 1, 2 and 3: Fig 1: 87pp: English.  
 XX  
 CC The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine proceeding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC yield a family of mRNAs each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor, ie. one or more of these exons may contain  
 CC sequences which direct proteolytic cleavage, adhesion, organization of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 XX  
 SQ Sequence 1404 AA:

Query Match 99.6%; Score 7245.5; DB 13; Length 1404;  
 Best Local Similarity 97.1%; Pred. No. 0;  
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAMTLPYIIIIIIIVIOVSSQ----- 25  
 DB 1 mawtclpylyllllsvtlvgvsgdlsacgrcgegysrdacncdyncqhyamecpdt 60  
 QY 26 -----ELSCGRGCFESFERGECDDAOCKKYDKCCPDVESCACVHNTPSPSSKAP 79  
 DB 61 kvtcaetlaeskyrcfcsfergrecddagckkydkccpdyescacvhnptpskapp 120  
 QY 80 PFGASGTTISTTKRSRPPNKKRKKVISEITEHSVSENOESSSSSSSSSSSTIM 139  
 DB 121 psgasgltksttkrppnkkkkkvlseeslteehsvenqesssssssssstliw 180

QY 140 KIKSSKNSAANRELQKTLVKONKKNRTKKKPTPPVVDAGSGDNGDKRYTTPNST 199  
 DB 181 ktkssksnaanrelqktlkvndknrtkkkppvvdagsgldngdkrtvtpdst 240  
 QY 200 TQHNKVTSPKITTAKPINRPSPSLPPNSDTSEKSLTVNKEETVEETTTNNKQTSIDG 259  
 DB 241 tqhnkvstspkittakpinrpsslpnsdtskelslvnketvctetcttnqtsldg 300  
 QY 260 KEKTSAAKESIOSIEKTSKDLAPTSKYLAKPTPAEITTKGPAALTTPKEPTTPPKBAS 319  
 DB 301 kektsaekesiosiektsakdlaptskylakpptaettkgpaalttpepttpkbas 360  
 QY 320 TTPKEPTTPITKSAPTTPKKEPAATTPPKSAPTTPKEPAATTPPKSAPTTPKEAPT 499  
 DB 361 ttpkepttpitksaptpkpepaattpkksaptpkpepaattpkksaptpk 540  
 QY 380 APPTTKSAPTTPKEPAATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPK 439  
 DB 421 apttksaptpkpepaattpkkapattppkkapattppkkapattppkkapattppk 480  
 QY 440 EPAPTAKPKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPAT 499  
 DB 481 epaptakpkapattppkkapattppkkapattppkkapattppkkapattppk 540  
 QY 500 TTKSAPTPKEPSPTTKKEPAATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPAT 559  
 DB 541 ttksaptpkepspttkpepaattpkkapattppkkapattppkkapattppkkapatt 600  
 QY 560 APAPKPKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTP 619  
 DB 601 apapkpapattppkkapattppkkapattppkkapattppkkapattppkkapatt 660  
 QY 620 PEPAPATTPKAAPNTPKKEPAATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPAT 679  
 DB 661 pepapattppkaapntpkkepaattpkkapattppkkapattppkkapattppkkapatt 720  
 QY 680 APPTPKKAPKELAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPKK 739  
 DB 721 aptpkpkapkelapattppkkapattppkkapattppkkapattppkkapattppk 780  
 QY 740 TAPPTKKEPAATTPPKKAPKELAPATTPPKKAPATTPPKKAPATTPPKKAPATTPPK 799  
 DB 781 tapptkkepaattpkkapkelapattppkkapattppkkapattppkkapattppk 840  
 QY 800 KPAATTPETPPPTSEVSTPTTKKEPTTHKSPDESPTLSAEPPTKALENSPKKEGVP 859  
 DB 841 kpaattpetppptsevspttkkeptthkspdesptlsaepppkalenspkkegvpt 900  
 QY 860 TTPPATKPEEMTTAKDKTTEDLKTTPETTTAAKPKETATTEKTESKITATTTQV 919  
 DB 901 ttppatkpeemttakdkttedlkttpetttaaakpketattekteskitatttqv 960  
 QY 920 TSTTTQDTTPPKITLTKTLTTLAPKVTYTTTKKTTTTTEINNKPEETAKPKDRATNSAATPK 979  
 DB 961 tstttqdtppkkitlktlktltaapkvtytttkkttttteinnkpeetakpkdratnsaatk 1020  
 QY 980 POKPTKAPKKTSTTKPKPTMPVRKPKTTPPKKATSTIMELNPTSRILAEAMLOTTPBN 1039  
 DB 1021 poptkapkktsttkpkptmpvrpktpppkkatstimelnptsrilaemlottpbn 1080  
 QY 1040 OTPNSKLVEVNPKESDAGAGGERPHMLRPHVMEPTVTPMDVYLPRVNOGIIINPMLS 1099  
 DB 1081 otpnsklvevnpkesdagaggrphmlrphvmeptvtpmdvylprvnoigiinpls 1140  
 QY 1100 DETNINCNRPVDTLTLNGTLVAFRGHYFMALSPFSPSPSARRTETVWGPSPTDIYFT 1159  
 DB 1141 detnincnrvdttlntglvafrghyfmalspfspspsarretetvwgpsptdiyft 1200  
 QY 1160 RCNCEGKTFEERKDSQWYFNTNDINDAGYKRPDIFKFGGGLIQOIAALSTAYKKMWPESVY 1219  
 DB 1201 rcncegkttfeerkdsqwyrntndindagyrpdkfkgggliqoiiaalstaykkmwpesvy 1260



sq : sequence 1404 AA;

[illegible]

Db 961 ttttqddtpfkittlktttlapkvtttkkitttelmkpeeta~~kk~~okdratnskattpk 1020



Region	Key	Location/Qualifiers
Region	Region	1..26
Region	Region	/label- Exon_I
Region	Region	26..67
Region	Region	/label- Exon_II
Region	Region	67..107
Region	Region	/label- Exon_III
Region	Region	107..157
Region	Region	/label- Exon_IV
Region	Region	157..200
Region	Region	/label- Exon_V
Region	Region	200..1141
Region	Region	/label- Exon_VI
Region	Region	1411..1166
Region	Region	/label- Exon_VII
Region	Region	1166..1212
Region	Region	/label- Exon_VIII
Region	Region	1213..1266
Region	Region	/label- Exon_IX



FT	Region	1266..1331
FT	Region	/label= Exon_X
FT	Region	1331..1373
FT	Region	/label= Exon_XI
FT	Region	1373..1404
FT	Region	/label= Exon_XII
XX		
PN	WO9213075-A.	
XX		
FD	06-AUG-1992.	
XX		
PF	17-JAN-1992;	92MO-US00433.
XX		
PR	18-JAN-1991;	91US-0643502.
PR	10-SEP-1991;	91US-0757022.
XX		
PA	(GEMT ) GENETICS INST INC.	
XX		
RI	Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;	
DR	WP1: 1992-284660/34.	
DR	N-PSDB: AA027223.	
XX		
PS	Claim 1, 2 and 3; Fig 1; 87pp; English.	
XX		
PT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
XX	bacterial and viral infections, etc.	
XX		
CC	The sequence given is a full length translation from the megakaryocyte	
CC	stimulating factor (MSF) precursor. The sequence covered by exons II,	
CC	III and IV encodes megakaryocyte stimulating factor (MSF). This	
CC	sequence is modified by the addition of an N-terminal sequence encoding	
CC	a secretory leader, an initiating methionine proceeding exon II and a	
CC	terminating codon following exon IV. The cDNA sequence given contains	
CC	sequences derived from human megakaryocyte colony stimulating factor	
CC	(meg-CSF). Exon I contains the initiating methionine, and encodes a	
CC	classical mammalian protein secretion signal sequence. The sequence	
CC	terminates in the region between amino acid residues 134 - 147. The	
CC	primary transcript of this gene may be cleaved in different ways to	
CC	yield a family of mRNAs each encoding a different MSF protein. Exons	
CC	II and VI are thought to be related to the activity of the factor and	
CC	are also implicated in the stability, folding and processing of the	
CC	molecule. These exons are also thought to play a role in the observed	
CC	synergy of MSF with other cytokines. Exons V - XII are believed to be	
CC	implicated in the processing or folding of the appropriate structure of	
CC	the resulting factor, ie. one or more of these exons may contain	
CC	sequences which direct proteolytic cleavage, adhesion, organisation of	
CC	the cellular matrix or extracellular matrix processing. Both naturally	
CC	occurring and non-naturally occurring MSF's may be characterised by	
CC	various combinations of alternatively spliced exons from this sequence,	
CC	with the exons spliced together in differing orders to form different	
CC	members of the MSF family.	
XX		
SQ	Sequence 1404 AA:	
QY	Query Match	98.7%; Score 6727; DB 13; Length 1404;
DB	Best Local Similarity	90.5%; Pred. No. 0;
QY	Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;	
QY	1 MARKTPIYLLLSVFIQOVSSQ-----	25
DB	1 maaktiplyllllsvfiqvasqdlssacagcsgesyrdatcncdynqhyamecpdf	60
QY	26 -----ELSCRGKCFSEFREGRECDADACKKTKCCPDYESFCAE-----	65
DB	61 kvtcaelsckgrcfesfergrecddacqkkydcpcpyesfcaeinvhmpscpskrap	120
QY	66 -----	65
DB	121 ppsagatstktkrspkpkpkkkkkkvleeseelteesvseengessssssssstliw	180

QY	66 -----VKDNKNKNTKKKPPPKPPVVDAGGGLDNGEFKVTTPDTST	106
DB	181 kkesksaanrelqklkxkdnkknckkppkpvvdaagsgldngdfkvtptst	240
QY	107 TOHNKYSTSKITTAKEINRPSLPPNSDTSKETSLVNEKTVERTTTTKQSTDG	166
DB	241 tqnhkvsstpkltakpindrpelppndstskstslvneketctellnkgtstdg	300
QY	167 KEKTSKAKETQSIEKTSADIASTVLAKEPPKAETTTGPALETTPKKEPTTPKEPAS	226
DB	301 kektsaketqslektssakdlaptskvlakpkaetctckgpalctkpkapttkpepas	360
QY	227 TTKKEPTPTTKSAPTTKKEPAATTTKSAATPKKEPAATTTKKEPAATTTKKEP	286
DB	361 tkpkeptpttkisapttkpkpkepaatctkspckpkpkepaatctkpkapttkkep	420
QY	287 APTTKSAPTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTK	346
DB	421 apttkksapttkpkpkepaatctkpkpkepaatctkpkapttkpkapttkpkapt	480
QY	347 EPAPPAKPKKPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKSAATTTKKEPA	406
DB	481 epapkapkpkapttkpkpkepaatctkpkpkepaatctkpkapttkksapttckep	540
QY	407 TTKSAPTTKKEPSPTTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKE	466
DB	541 tkksapttkpkpspttkpkpkepaatctkpkpkepaatctkpkapttkpkapttkpk	600
QY	467 APAPKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA	600
DB	601 aptapkepaatctkpkpkepaatctkpkpkepaatctkpkapttkpkapttkpk	660
QY	527 PEEAPATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKE	586
DB	661 peepaptpkpaapntpkpkepaatctkpkpkepaatctkpkapttkpkapttkpk	720
QY	587 APTTKKPAKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKE	646
DB	721 apttkpkpkepaatctkpkpkepaatctkpkpkepaatctkpkapttkpkapttk	780
QY	647 TAPTTLKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPAATTTKKEPA	706
DB	781 tapttlkepaatctkpkpkepaatctkpkpkepaatctkpkapttkpkapttkpk	840
QY	707 KPAPTTPETPPPTTSVSPPTTKKEPTTHKSPDSETPELSAEPPKALENPKRGPVT	766
DB	841 kpapttpepppttsvsppttkpkpkepaatctkpkpkepaatctkpkapttkpk	900
QY	767 TTPPAATKPEMTTAKDKTTEHDLTRTPETTTAAKMKETATTTKTESKITAATTTQV	826
DB	901 tktpaatkpemttakkdttedtrtpetcttaapkmkkaetttkteskittattqv	960
QY	827 TSTTQODTTPRKITTLKTTTTLAPVTTTKKTITTEINMKPEETAKPRDRATNKATPK	886
DB	961 tsttgdtctpklitlktlktlktlktlktlktlktlktlktlktlktlktlktlktl	1020
QY	887 POKTPAKPKKPTSTKPKTPRVRKPKTPPKRGTSTMPELNPTSIAMLOTTTRPN	946
DB	1021 pqktpakpkpkptstskpktpvrkpktpckpckpckpckpckpckpckpckpckp	1080
QY	947 QTPNSKLVNPKSEDAGAGBETPHMLLRPHFMPEVTDDMOYLPRVPRNOGIINPMS	1006
DB	1081 qtpnslvsnpkseagagsetphmlrrphfmpevtddmoyleprvpngslllppms	1140
QY	1007 DEINICNGKPVVDGTLTARNITLAFRGHYFMALSPSPSPARRITVEWGISPIDTFT	1066
DB	1141 deinincgkpvvgdgtlltarnitlvafrghyfmalspspsparritvewgispidtft	1200
QY	1067 RCNCEGKTEFFFKSOYWRFTNDIKDAGYPKPIFKGSGGLTGQIVAALESTAKKMPESY	1126
DB	1201 rcncegkttfffksgywrftndikdegypkplfkfglgtgivaalestakynpessy	1260





Query Match	98.7%	Score 6727;	DB 22;	Length 1404;
Best Local Similarity	90.5%;	Pred. No. 0;		
Matches 1970; Conservative	0;	Mismatches	0.	Indels 134.

Qy	1	MAKTLPIYLLLLSVFYIQQVSSO-----	2
Dd	1	maaktpiylillllsvfiyqvsqdlssacagrcgeysrdatcnodnqhymeccpdt	6
Qy	26	-----ELSCGRCESEFERGECDOACQKXVKKCCPDVESCFAE-----	65
Dd	61	krvtcaeljschgrcfesfergrecodacqckkydkccpdyefcaehvnltpssakkap	12
Qy	66	-----	65
Dd	121	ppsaqclskstkrspkpnkkkkkviaseelthehsvengeassssssssasltw	18
Qy	66	-----VQDNKKNNTKKKKPKPPKPPVDEAGSLGNDGFRVYTPDST	10
Dd	181	k1ksqksnaaarelqkklxkxdkmknkkkpkpdpvdeaagsgldngdfkvtptdst	24
Qy	107	TOHKNVTSPIITTAKPINRPSLPKNSDPSKETSPLVKNKETTVEKETTNNKQSTDGC	16
Dd	241	tqmhkvstspklttakpnlprslpnpsdstskststnkvctvelkettlnkclstsg	30
Qy	167	KEKTSAKENOSJEKTSAKOLAPTSVYLAKPVPKAAETTTKGPALTTPKKEPTTPKEKBPAS	22
Dd	301	kektsaekueqslaktsakalapskvlaklpkaeettckgalaltkpkepttpkpeas	36
Qy	227	TTPKKEPTPIITKSAFTTPKBPATTTKSAFTTPKBPATTTKEBPATTTKEBPATTTKEP	28
Dd	361	tkpkepttkksaepcttkpkkeapcttkksapcttkpeapcttkpeapcttkpeapcttkpe	42
Qy	287	APTITKSAFTTPKEBPATTTPKKPAPTTPKBPATTTPKBPATTTPKKEPTTPKEBPATTPK	34
Dd	421	apcttkksaepcttkpeapcttkpkpapttkpeapcttkpepttkpeapcttkpeapcttkp	48
Qy	347	EPAPTAPEKKAPATTPKEBPATTPPKKEBPATTTKEBPATTTKEBPATTTKSAFTTTKEBAPT	40
Dd	481	epaptpakpkaprttkpeapcttkpkpeapcttkpeapcttkpeapcttkksapcttkkeapct	54
Qy	407	TTKTSAPTTKRESPTTTKBPATTTPKKEBAPTTPKKKAPATTPKEBPATTPKKEBPATTTTKP	46
Dd	541	ctksaepcttkpeapcttkkeapcttkpkpeapcttkpkpapttkpeapcttkpeapcttkkp	60
Qy	467	APAPAEBPATTPPKETAPTTPKKLTAPTTPEKLAPTTPEKAPATTPPEELAPTTPEEPTPT	52
Dd	601	apapaepepcttkpetapcttkpkltapcttkpeklapcttkpeapcttkpeelapcttkpeepct	66
Qy	527	PEBPATTTKAAAPNTPKBPATTPPKBPATTPPKBPATTPPKBPATTPPKGTAAPTLKPE	58
Dd	661	peebpapttkaaapntpkpeapcttkpkpeapcttkpkpeapcttkpketapcttkpqlapcttkkep	72
Qy	587	APTTPKPKAPKELATTTKEPTSTTSOKAPATTPPKGTAAPTTKEBPATTPKBPATTPPKG	64
Dd	721	aptpkpkapkelapcttkkeptsttsckspcttkpqtapcttkpeapcttkpeapcttkpqt	78
Qy	647	TAPTLKEBPATTPKPKAPKELATTTKGPSTTSOKAPATTTKEBPATTTKEBPATTTK	70
Dd	781	taptlkeapcttkpkpapelapcttkqplsttsckapcttkpeapcttkpeapcttkp	84
Qy	707	KPAPTTPPTPTTSVSTPTTTKEPTTHKSPDNESTPELSAEPPTKALENSPKRPGVPT	76
Dd	841	kpaepcttkpplpttsvstpttkcttkcpttklnkspdesptelsaepkpalenspkrgvpt	90
Qy	767	TKTPAPATPEWTTAKKNTTERDLRTPTPTNTTAAPKMKETAATTEKTTESKITAATTTQV	82
Dd	901	tktpaepcttkenttakdctterdlrtptptnttaapmketaattekteskitaatqtvy	96
Qy	827	TSTTQDTPPEFKITTLAKTTTLAPKVTTKKTTTTTEINMKPEETAKPKDRTANSKATTPK	88
Dd	961	tstctqcttpcfikltlkttlapkvttcttkltlteimkpeetakpkdrtanskatctpk	102



DR N-PSDB: AAR98981.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use  
 XX  
 PS Claim 20; Page 1198-1201; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC biophysics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 SO Sequence 1299 AA

Query Match 100.0%; Score 5011; DB 22; Length 1299;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-273;  
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDNKRNKRRKPPKPPVNDAGSLNDGDKVTPDSTQHNKVSPTKTKAKIN 60  
 DB 20 VDNKRNKRRKPPKPPVNDAGSLNDGDKVTPDSTQHNKVSPTKTKAKIN 259  
 QY 61 PRPSLPNSDTSKETSLSLVNKTVEKETTNNKQSTDKGKETSKEKTSK 120  
 DB 60 PRPSLPNSDTSKETSLSLVNKTVEKETTNNKQSTDKGKETSKEKTSK 319  
 QY 121 DLAPTSKTLAPPPKAEPTTGAPATTPKPEPTTPPKPASTTPKPKPA 180  
 DB 120 DLAPTSKTLAPPPKAEPTTGAPATTPKPEPTTPPKPASTTPKPKPA 379  
 QY 181 EPAPTSKATTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 240  
 DB 180 EPAPTSKATTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 439  
 QY 241 PKKPAPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 300  
 DB 240 PKKPAPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 499  
 QY 440 PKKPAPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 360  
 DB 440 PKKPAPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 559  
 QY 500 PTPKPEPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 420  
 DB 500 PTPKPEPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 619  
 QY 421 TPKKLPTPEKLAPTPEKPAPTTPKPASTTPKPKPASTTPKPKPA 480  
 DB 420 TPKKLPTPEKLAPTPEKPAPTTPKPASTTPKPKPASTTPKPKPA 679  
 QY 620 TPKKLPTPEKLAPTPEKPAPTTPKPASTTPKPKPASTTPKPKPA 540  
 DB 620 TPKKLPTPEKLAPTPEKPAPTTPKPASTTPKPKPASTTPKPKPA 739  
 QY 481 PAPTPKPEPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 600  
 DB 480 PAPTPKPEPTTPKPEPTTPPKPASTTPKPKPASTTPKPKPASTTPK 799  
 QY 541 EPTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPASTTPKPKPA 660  
 DB 540 EPTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPASTTPKPKPA 759  
 QY 601 KELAPTTKPGTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPA 660  
 DB 600 KELAPTTKPGTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPA 859  
 QY 800 KELAPTTKPGTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPA 720  
 DB 800 KELAPTTKPGTSTSDKPAPTTPKPASTTPKPKPASTTPKPKPA 919  
 QY 661 PTTKEPTTHKSPDSTPELSAFTPKALNSKKEGVTPTTPATKEMTTAKDKT 780  
 DB 660 PTTKEPTTHKSPDSTPELSAFTPKALNSKKEGVTPTTPATKEMTTAKDKT 780  
 QY 721 TERDLRTPEPTTAAPKMTKETATTEKTESKITATTTQVSTTODTTPK 780  
 DB 720 TERDLRTPEPTTAAPKMTKETATTEKTESKITATTTQVSTTODTTPK 780

DB 920 terdlrtpepttaapnkctetatttektetklatitqvtsttqdtptkittlkt 979  
 QY 791 TLAPKVTTKKTTTTEINRPEETAKPRDRATNSKATTPKPKPTSTKPKPT 840  
 DB 980 TLAPKVTTKKTTTTEINRPEETAKPRDRATNSKATTPKPKPTSTKPKPT 1039  
 QY 841 MPRVAKPTTPPKKMTSTWELNPTSRISAMLOTTTRPQNSKLVEVNPKEADAG 900  
 DB 1040 mprvaxkpttpkkmstwpelnptsrissamlootttrpqnskllvevnpkedsdg 1099  
 QY 901 AEGEPHLLRPHVMPKPRVNDMDYLRVNPQGIINMWS 941  
 DB 1100 aegephlmlrphvmpkprvndmdylrvnpqgillnplms 1140

RESULT 2  
 AAR26049  
 ID AAR26049 standard; Protein: 1404 AA.  
 AC AAR26049;  
 XX  
 DT 02-FEB-1993 (first entry)  
 XX  
 DE MSF precursor.  
 XX  
 KM Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
 KM stability; proteolytic cleavage; adhesion; alternative splicing.  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Region 1..26  
 FT /label= Exon\_I  
 FT Region 26..67  
 FT /label= Exon\_II  
 FT Region 67..107  
 FT /label= Exon\_III  
 FT Region 107..157  
 FT /label= Exon\_IV  
 FT Region 157..200  
 FT /label= Exon\_V  
 FT Region 200..1141  
 FT /label= Exon\_VI  
 FT Region 1411..1166  
 FT /label= Exon\_VII  
 FT Region 1166..1212  
 FT /label= Exon\_VIII  
 FT Region 1213..1266  
 FT /label= Exon\_IX  
 FT Region 1266..1331  
 FT /label= Exon\_X  
 FT Region 1331..1373  
 FT /label= Exon\_XI  
 FT Region 1373..1404  
 FT /label= Exon\_XII  
 XX  
 PN MO9213075-A.  
 XX  
 PD 06-AUG-1992.  
 XX  
 PF 17-JAN-1992; 92MO-US00433.  
 XX  
 PR 18-JAN-1991; 91US-0643502.  
 PR 10-SEP-1991; 91US-0757022.  
 XX  
 PA (GENE) GENETICS INST INC.  
 XX  
 PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
 DR WPI; 1992-284660/34.  
 DR N-PSDB; AA027223.  
 XX  
 PT New human mega-karyocyte stimulating factors - for treating



PT Immune deficiencies, cancer, exposure to radiation or drugs,  
 PT bacterial and viral infections, etc.  
 XX  
 PS Clalm 1, 2 and 3; Fig 1; 87pp; English.  
 CC The sequence given is a full length translation from the megakaryocyte  
 CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
 CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
 CC sequence is modified by the addition of an N-terminal sequence encoding  
 CC a secretory leader, an initiating methionine preceding exon II and a  
 CC terminating codon following exon IV. The cDNA sequence given contains  
 CC sequences derived from human megakaryocyte colony stimulating factor  
 CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
 CC classical mammalian protein secretion signal sequence. The sequence  
 CC encoding the original meg-CSF includes exons II-IV and is thought to  
 CC terminate in the region between amino acid residues 134 - 147. The  
 CC primary transcript of this gene may be cleaved in different ways to  
 CC yield a family of mRNA's each encoding a different MSF protein. Exons  
 CC V and VI are thought to be related to the activity of the factor and  
 CC are also implicated in the stability, folding and processing of the  
 CC molecule. These exons are also thought to play a role in the observed  
 CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
 CC implicated in the processing or folding of the appropriate structure of  
 CC the resulting factor, i.e. one or more of these exons may contain  
 CC sequences which direct proteolytic cleavage, adhesion, organization of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 XX  
 XX  
 Sequence 1404 AA:

Query Match 100.0%; Score 5011; DB 13; Length 1404;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-273;  
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNDNKNRTKRRKPPRVVDEAGSLDNGDFVTPDSTTQHNKVSPPKTKAPIN 60  
 DB 200 VKDNKNRTKRRKPPRVVDEAGSLDNGDFVTPDSTTQHNKVSPPKTKAPIN 259  
 QY 61 PRPSLPNSDTSKETSIVNKKETVETKTTNNKQTSDEKETSKEQSIKTSK 120  
 DB 260 PRPSLPNSDTSKETSIVNKKETVETKTTNNKQTSDEKETSKEQSIKTSK 319  
 QY 121 DLAPTSKVLAKPTPKAETTKGPALETPKPEPTTPKPEASTPKPEPTTKSA 180  
 DB 320 DLAPTSKVLAKPTPKAETTKGPALETPKPEPTTPKPEASTPKPEPTTKSA 379  
 QY 181 EPAPTTTSAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 240  
 DB 380 EPAPTTTSAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 439  
 QY 241 PKKPAPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 300  
 DB 440 PKKPAPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPE 499  
 QY 341 PTPPEAPPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 360  
 DB 500 PTPPEAPPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP 559  
 QY 361 PAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 420  
 DB 560 PAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 619  
 QY 421 TPKKLTPTPKELAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 480  
 DB 620 TPKKLTPTPKELAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 679  
 QY 481 PAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 540  
 DB 680 PAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 739

QY 541 EPTSTSDKAPPTPKGAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 600  
 DB 740 EPTSTSDKAPPTPKGAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 799  
 QY 601 KELAPPTTKGPTSTSDKAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 660  
 DB 800 KELAPPTTKGPTSTSDKAPPTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 859  
 QY 661 PTTTEPTTKRSPESPEPELSAETTPALNSPKPEVPTTKTPAATKPMATK 720  
 DB 860 PTTTEPTTKRSPESPEPELSAETTPALNSPKPEVPTTKTPAATKPMATK 919  
 QY 721 TERDRTPTPTTAAPKKTETATTEKTSKITATTOYSTTODTTPFKITLTKT 780  
 DB 920 TERDRTPTPTTAAPKKTETATTEKTSKITATTOYSTTODTTPFKITLTKT 979  
 QY 781 TLAPVTTTKKTTTTEITMKNRPEETAKRKDRATNSKATTPPKPKPKPTSTK 840  
 DB 980 TLAPVTTTKKTTTTEITMKNRPEETAKRKDRATNSKATTPPKPKPKPTSTK 1039  
 QY 841 MPVKKPPTPTPKMTSTMPLELNPTRISANLQTTTRPQNTNSKLVENPKSEDAG 900  
 DB 1040 MPVKKPPTPTPKMTSTMPLELNPTRISANLQTTTRPQNTNSKLVENPKSEDAG 1099  
 QY 901 AEGETPHMLRPHVFMPEVTPDMOYLPRVPMOQIIRPMLS 941  
 DB 1100 AEGETPHMLRPHVFMPEVTPDMOYLPRVPMOQIIRPMLS 1140

RESULT 3  
 AAB60568  
 ID AAB60568 standard; Protein; 1404 AA.

AC AAB60568;  
 XX  
 XX 27 APR-2001 (first entry)  
 DE Human megakaryocyte stimulating factor (MSF, CACP).  
 XX  
 XX Human: CACP protein; camplocactylly-arthropathy-coxa vara-pericarditis;  
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 KW antidiabetic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200107068-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 XX 21-JUL-2000; 2000MO-US20002.  
 PF  
 XX 23-JUL-1999; 99US-0145328.  
 PR 19-JUL-2000; 2000US-0145328.  
 XX  
 PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX  
 PI Warman ML;  
 XX  
 DR WPI; 2001-182721/18.  
 XX  
 PT New composition comprising the camplocactylly-arthropathy-coxa  
 PT var-pericarditis protein in combination with an anesthetic, useful for  
 PT treating osteoarthritis, or as lubricants of tissue and joints  
 PS Example 1; Page 3; 34pp; English  
 XX  
 CC The invention relates to a method of treating osteoarthritis via the  
 CC administration of a composition comprising the camplocactylly-arthropathy-  
 CC coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
 CC The composition may further comprise a local anesthetic. The composition  
 CC of the invention may be administered via intra-articular or intravenous



SE A 1D NO: 1, 200-1167  
AC NO: AAR26049, A-gensay-1101  
File Copy  
Page 219

DR N-PSDB: AAR98981.

Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use

Claim 20: Page 1198-1201: 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

Sequence 1299 AA:

Query Match 100.0%; Score 5155; DB 22; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 3.9e-282;  
Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VKDNKKNTKKKPPKPPVVDAGSGLDNGDFKVTPTSTTOHNVSTSPKITTAKPIN 60  
200 VKDNKKNTKKKPPKPPVVDAGSGLDNGDFKVTPTSTTOHNVSTSPKITTAKPIN 259  
61 PRSLPFRNSDTSKESLTVNKETTVEKETTNNKOSTDGEKETSARETOSIEKTSAK 120  
260 PRSLPFRNSDTSKESLTVNKETTVEKETTNNKOSTDGEKETSARETOSIEKTSAK 319  
121 DLATSKIAKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 180  
320 DLATSKIAKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 379  
181 EPAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 240  
380 EPAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 439  
241 PKRAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 300  
440 PKRAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 499  
301 PTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 360  
500 PTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 559  
361 PAPTTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 420  
560 PAPTTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 619  
421 TPKKLPTTKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 480  
620 TPKKLPTTKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 679  
481 PAPTTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 540  
680 PAPTTPKREAPPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 739  
541 EPSTISDRAPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 600  
740 EPSTISDRAPTTKSAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 799  
601 KELAPPTTKGPTSTSKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 660  
800 KELAPPTTKGPTSTSKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 859  
661 PTTTKREPTTKHKSPPDESTPELSAEPKALANSKPEKAVTTTKPAATKREMTTAKDKT 720  
860 PTTTKREPTTKHKSPPDESTPELSAEPKALANSKPEKAVTTTKPAATKREMTTAKDKT 919  
721 TERDLPTTPTTAAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 780

Db 920 terdlpttptttaaapkmktetatttkttskltatttqvtstttdgtlftlitt 979  
Qy 781 TLAPKVTYTTKTTTTEIMNKEETANKPRDRATNSKATTPKPKAPKPTSKRPKT 840  
Db 980 TLAPKVTYTTKTTTTEIMNKEETANKPRDRATNSKATTPKPKAPKPTSKRPKT 1039  
Qy 841 MPVRKRTTTPKRNKSTNPELNPSTRIABAMLOTTTRPNOTPNSKIVEVNPNSKEDAGG 900  
Db 1040 MPVRKRTTTPKRNKSTNPELNPSTRIABAMLOTTTRPNOTPNSKIVEVNPNSKEDAGG 1099  
Qy 901 AEGEPHMLPHEVPEVPEVMDQYLDPRVPMOGIINPMUSDENICNGKRVDDGLTIRN 960  
Db 1100 AEGEPHMLPHEVPEVPEVMDQYLDPRVPMOGIINPMUSDENICNGKRVDDGLTIRN 1159  
Qy 961 GTLVAFRG 968  
Db 1160 gtlvafrg 1167

RESULT 2

AAR26049 ID AAR26049 standard; Protein: 1404 AA.

AC AAR26049;

DT 02-FEB-1993 (first entry)

DE MSF precursor.

KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF; stability; proteolytic cleavage; adhesion; alternative splicing.

OS Synthetic.

FT Key Location/Qualifiers

FT 1..26 /label= Exon\_I

FT 26..67 /label= Exon\_II

FT 67..107 /label= Exon\_III

FT 107..157 /label= Exon\_IV

FT 157..200 /label= Exon\_V

FT 200..1141 /label= Exon\_VI

FT 1141..1166 /label= Exon\_VII

FT 1166..1212 /label= Exon\_VIII

FT 1212..1266 /label= Exon\_IX

FT 1266..1331 /label= Exon\_X

FT 1331..1373 /label= Exon\_XI

FT 1373..1404 /label= Exon\_XII

PN MO9213075-A.

PD 06-AUG-1992.

PF 17-JAN-1992; 92MO-US00433.

PR 18-JAN-1991; 91US-0643502.

PR 10-SEP-1991; 91US-0757022.

PA (GENE) GENETICS INSTR INC.

PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;









FILE COPY  
AC: NO: AAR26049, A-GeneSeq-1104  
Page 2

DR NPSDB: AAR98981.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use

PS Clalim 20; Page 1198-1201; 1275pp; English.

CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cqa, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC fidenatics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.

XX Sequence 1299 AA:

Query Match 100.0%; Score 5416; DB 22; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 2.4e-298;  
Matches 1013; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRTKRPKPPVVDAGSGLDNGDFKVTPTDSTTOHNVSTSPKTTAKPIN 60  
DB 200 VKDNKKNRTKRPKPPVVDAGSGLDNGDFKVTPTDSTTOHNVSTSPKTTAKPIN 259  
QY 61 PRSLPRSDTSKETSLSLVNKKETVETKETTNNKQSTDEKKTSAKQSTIEKTSK 120  
DB 260 PRSLPRSDTSKETSLSLVNKKETVETKETTNNKQSTDEKKTSAKQSTIEKTSK 319  
QY 121 DLAPTSKVLAKPKAEKTTKAPLTPPKKPTTPPKKPAATTPPKKPAATTPPK 180  
DB 320 DLAPTSKVLAKPKAEKTTKAPLTPPKKPTTPPKKPAATTPPKKPAATTPPK 379  
QY 181 EPAPTTKSAPTPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPK 240  
DB 380 EPAPTTKSAPTPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPK 439  
QY 241 PKRPAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 300  
DB 440 PKRPAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 499  
QY 301 PTPKPEAPATTPKPSPTPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 360  
DB 500 PTPKPEAPATTPKPSPTPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 559  
QY 361 PAPTPKPEAPATTPKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPK 420  
DB 560 PAPTPKPEAPATTPKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPK 619  
QY 421 TPKKLTPTTPPKLAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 480  
DB 620 TPKKLTPTTPPKLAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPA 679  
QY 481 PAPTPKPEAPATTPKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPK 540  
DB 680 PAPTPKPEAPATTPKPAATTPPKKPAATTPPKKPAATTPPKKPAATTPK 739  
QY 541 EPNTSTSDKAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTP 600  
DB 740 EPNTSTSDKAPTPPKKPAATTPPKKPAATTPPKKPAATTPPKKPAATTP 799  
QY 601 KELAPTTKGTSTSTSDKAPTPPKKPAATTPPKKPAATTPPKKPAATTP 660  
DB 800 KELAPTTKGTSTSTSDKAPTPPKKPAATTPPKKPAATTPPKKPAATTP 859  
QY 661 PTTTKEPTTHKSPDESTPELSAEPTRKALENSEKPEGVPTTKPAATPKEMTTAKDKT 720  
DB 860 PTTTKEPTTHKSPDESTPELSAEPTRKALENSEKPEGVPTTKPAATPKEMTTAKDKT 919  
QY 121 TEHDLRTPETTTAARPMETATTTTETTESKLTATTTTVOVSTTTODTTPKITTAKTT 760  
DB 121 TEHDLRTPETTTAARPMETATTTTETTESKLTATTTTVOVSTTTODTTPKITTAKTT 760

DB 920 TELRTTPETTTAARPMETATTTTETTESKLTATTTTVOVSTTTODTTPKITTAKTT 979  
QY 781 TLAPRTTTRKRTTTRTTEINMKPETAKPKDRATNSKATTPKPKPTAKPKPTSKRKT 840  
DB 980 TLAPRTTTRKRTTTRTTEINMKPETAKPKDRATNSKATTPKPKPTAKPKPTSKRKT 1039  
QY 841 MPVRKRTTTPPKKRTSTNPELNPSTRIAEAMQTTTREPNTGNSKLVENPKSEDAG 900  
DB 1040 MPVRKRTTTPPKKRTSTNPELNPSTRIAEAMQTTTREPNTGNSKLVENPKSEDAG 1099  
QY 901 AEGTPEMLLRPVHMEVPPDDYLPBPVNOGIIINPMISDETNICNGKPVNGITTLRN 960  
DB 1100 AEGTPEMLLRPVHMEVPPDDYLPBPVNOGIIINPMISDETNICNGKPVNGITTLRN 1159  
QY 961 GTVAFRGHYFMMLSPPSPSPAPNRITTEVWGISPIDVTFRNCDEGKTFEFK 1013  
DB 1160 GTVAFRGHYFMMLSPPSPSPAPNRITTEVWGISPIDVTFRNCDEGKTFEFK 1212

RESULT 2

AAR26049 standard; Protein; 1404 AA.

AC AAR26049;

DT 02-FEB-1993 (first entry)

DE MSF precursor.

KM Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
KW stability; proteolytic cleavage; adhesion; alternative splicing.

OS Synthetic.

XX Key Location/Qualifiers

FT Region 1..26

FT Region /label= Exon\_I

FT Region 26..67

FT Region /label= Exon\_II

FT Region 67..107

FT Region /label= Exon\_III

FT Region 107..157

FT Region /label= Exon\_IV

FT Region 157..200

FT Region /label= Exon\_V

FT Region 200..1141

FT Region /label= Exon\_VI

FT Region 1411..1166

FT Region /label= Exon\_VII

FT Region 1166..1212

FT Region /label= Exon\_VIII

FT Region 1213..1266

FT Region /label= Exon\_IX

FT Region 1266..1331

FT Region /label= Exon\_X

FT Region 1331..1373

FT Region /label= Exon\_XI

FT Region 1373..1404

FT Region /label= Exon\_XII

XX MO9213075-A.

XX PD 06-AUG-1992.

XX PF 17-JAN-1992; 92MO-US00433.

XX PR 18-JAN-1991; 91US-0643502.

XX PA 10-SEP-1991; 91US-0757022.

XX PI (GENE) GENETICS INST INC.

XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;



DR WPI; 1992-284660/34  
DR N-P; DB; AAQ27223.  
XX

new human mega-karyocyte stimulating factors - for treating immune deficiencies, cancer, exposure to radiation or drugs bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Flg 1; 87pp; English.  
XX

The sequence given is a full length translation from the megakaryocyte stimulating factor (MSF) precursor. The sequence covered by exons II, III and IV encodes megakaryocyte stimulating factor (MSF). This sequence is modified by the addition of an N-terminal sequence encoding a secretory leader, an initiating methionine preceding exon II and a terminating codon following exon IV. The cDNA sequence given contains sequences derived from human megakaryocyte colony stimulating factor (meg-CSF). Exon I contains the initiating methionine, and encodes a classical mammalian protein secretion signal sequence. The sequence encoding the original meg-CSF includes exons II-IV and is thought to terminate in the region between amino acid residues 134 - 147. The primary transcript of this gene may be cleaved in different ways to yield a family of mRNAs, each encoding a different MSF protein. Exons IV and V are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, i.e. one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSFs may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family.

Sequence	1404 AA;
50	

Query Match	100.0%;	Score 5416;	DB 13;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 2.6e-298;		
Matches 1013;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	VKONKNNRKKRPKRPVNVDAAGSLDNODFKYTPEDSTOHNKVNYSRKTAKPLN	60
Db	200	vndknknrtkkkpprkpprvvdaagsjldnglfnfkrcpdrctclnknvysaspklttakp	25
QY	61	PRPSLEPNSDISKETSLSLVNKEETTVEKRTTTTTNKKQSTDGKEKTSIAKETOSIEKTSAK	12
Db	260	pprsjlpnsdtketslslyvnkettvekkettlnkgtsldgkettssaketsjlekttsak	31
QY	121	DLAPTSKVLAKRTPAEITTTGSLALTTTPKEPTPTTPKEBASSTTPKKEPTTTTISAPTPPK	180
Db	320	dlaptskvlakrcpbaeettcglpaltltpkcpctlpkpeasttpkcpctlttkksapctpk	37
QY	181	EPAPTTTTKSAPTTKEBAPTTTTKEBAPTTTKEBAPTTTKEBAPTTTTKSAPTTPKBAPPT	24
Db	380	epapcttksapctlpkpeapcttkkeapctlpkpeapcttkkeapcttkksapctpkpeapct	43
QY	241	PKKRPAPTTPKBAPPTPKREPTPTTPPKBAPTTTKEBAPTTPKBAPPTPKBAPPTPKBPA	300
Db	440	pkkrpapttkpeapctlpkpeapcttkkeapcttkkeapcttkkeapcttkkeapcttkkpa	49
QY	301	PTTPKEBAPTTTKEBSPPTPKBAPTTTTKSAPTTKEBAPTTTTKSAPTTPKRBSPTTTKE	360
Db	500	pttpkeapcttkkebspptpkpeapcttkkeapcttkkeapcttkkeapcttkkeapcttkke	55
QY	361	PAPPTPKBAPATTTPKKRAPTTTPKBAPTTPKBAPTTTTKKRAPTAAPKBAPTTPKETAPT	420
Db	560	paptpkpeapcttkpkpapttkpkpeapcttkpkpeapcttkpkpapttkpkpapttkpketa	61
QY	421	TPKTLPTTPPKLAPTTPEKRPAPTTPEELAPTTPEEPPTTPEEBAPTTTPRAAANPPEKE	480
Db	620	tpkrltpttpkclapttpekrapttpeelapttpeeppttkpeebapcttkpraanpke	67

QY	481	PAPTPKCEBAPTTPKEBPAPTTPKXETAAPTTPKGAAPTTPLEKEBAPTTPKRAAPKELAPTTPK	540
Db	680	papctcpkpaaptcpkpeapctcpkacaptcpkxgaptcpkilekpaaptcpkpkapkelaptcpk	739
QY	541	EPTSTTSOKPAPTTPKGTAPTTPKEBAPTTPKREBAPTTPKGTAPTLLEBAPTTPKPKAP	600
Db	740	epstctsdkapctcpkxgaptcpkpeapctcpkpaaptcpkxgaptcpkilekpaaptcpkpkap	799
QY	601	KELAPTTPKGTSTTSOKPAPTTPKELAPPTTPKEBAPTTPKRAAPTTPETPPPTTSSEVT	660
Db	800	kelapctcpkxgptctsdkapctcpkpaaptcpkpaaptcpkpkapctcpkpaaptcpkpkap	859
QY	661	PTTKEPTTHKSDSESTPELSAEPPTKALENSPPKEBGPVTTKPAATKPEMTTAKDKT	720
Db	860	pttkeptthkspdeestpelsaepptkalsenpkapctcpkpaaptcpkpaaptcpkpaapt	919
QY	721	TERDRTPEPTTAPAPKMTKETAATTPTEKTESKITAATTPQVSTTQDTPPKTITTLKTT	780
Db	920	terdrttpepttapaapmketatactctekteskilectcpqsttqddctcpkltlkt	979
QY	781	TLAPKVTTKTITTTBINKPPEETAKPKDRAINSKATTPKOPKPAKPKSTPKRPT	840
Db	980	tlapkvttktlcttlttlimkpeetakpkdratnskatcpkqpkpkapckpctckpct	1039
QY	841	MPAPRKPTTPTPPKMTSTMBELNPTPSRTAEALQTTTPRNOTPNKSLVEVNPKSEBAG	900
Db	1040	mpaprkpttptpckmtstmbelnptpsrtaealqtttprnotpnkslvevnpksebag	1099
QY	901	AEGSTPMALLRPHVEMPEVDPMDYLPBPVNOGIIINPMALSDETNINCNGKPDVGLTTLRN	960
Db	1100	aegstpmallrphvfmpevdpmdylpbpvngiiinpmalsetnlcnkgkpdvglttlrn	1159
QY	961	GTLVAFRGHFWMLSPSPSPAPARRITTEWGLPSPIDIYFTNCNGBGKTFEPK	1013
Db	1160	gtlvafrgfhfwmlspfspspaparrittewglpspidiyftncnbgktfepk	1212

RESULT	3
AAB60568	
ID	AAB60568 standard; Protein; 1404 AA
XY	

AA	AB60568;	
XX		
XX		
DT	27-APR-2001 (first entry)	
XX		
XX	Human megakaryocyte stimulating factor (MSF, CACP).	
XX		
XX	Human: CACP protein; campitodactylly-arthropathy-coxa vara-pericarditis;	
XX	MSF; megakaryocyte stimulating factor; synovial lubricant;	
XX	chromosome 1q425-31; osteoarthritis; joint lubrication; osteopathic;	
KW	antiarthritic.	
XX		
OS	Homo sapiens.	
XX		
PN	NC0200107068-AL	
XX		
PD	01-FEB-2001.	
XX		
PF	21-JUL-2000; 2000MO-US20002.	
XX		
PR	23-JUL-1999; 99US-0145328.	
XX		
PR	19-JUL-2000; 2000US-0145328.	
XX		
PA	(UYCA-) UNIV CASE WESTERN RESERVE.	
XX		
PI	Warman ML;	
XX		
DR	WPI; 2001-182721/18.	
XX		
PT	Vara composition comprising the campitodactylly-arthropathy-coxa	
PT	vara-pericarditis protein in combination with an anesthetic, useful for	
PT	treating osteoarthritis, or as lubricants of tissue and joints -	









RA (GENE ) GENETICS INST INC.

PI Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;

XX WPI: 1992-284660/34.

DR N-PSDB: MA027223.

PT New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
XX bacterial and viral infections, etc.

PS Claim 1, 2 and 3; Fig 1; 87pp; English.

CC The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine preceding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNA's each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, i.e. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organization of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterized by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.

XX Sequence 1404 AA:

Query Match 100.0%; Score 5698; DB 13; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEDNKKRRTKRRTPPYVDAGSGIDNGDFKVTTPDSTOHKNVSRPKITAKPLN 60  
DB 200 VEDNKKRRTKRRTPPYVDAGSGIDNGDFKVTTPDSTOHKNVSRPKITAKPLN 60  
QY 61 PRPSLPNSDTSKETSIVNKKETVETKETTNNKOTSDGKETTSAKQSIKTSK 120  
DB 260 PRPSLPNSDTSKETSIVNKKETVETKETTNNKOTSDGKETTSAKQSIKTSK 120  
QY 121 DLAPTSKYLAKRTPKRAETTTGSPALTTPKEPTTPPKPASTTPKEPTTTKSAPTTPK 180  
DB 320 DLAPTSKYLAKRTPKRAETTTGSPALTTPKEPTTPPKPASTTPKEPTTTKSAPTTPK 180  
QY 330 DLAPTSKYLAKRTPKRAETTTGSPALTTPKEPTTPPKPASTTPKEPTTTKSAPTTPK 180  
DB 330 DLAPTSKYLAKRTPKRAETTTGSPALTTPKEPTTPPKPASTTPKEPTTTKSAPTTPK 180  
QY 131 EPAPPTTSAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTP 240  
DB 380 EPAPPTTSAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTP 240  
QY 241 PKKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPA 300  
DB 440 PKKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPA 300  
QY 301 PTPREPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 360  
DB 500 PTPREPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 360  
QY 361 PAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTP 420  
DB 560 PAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTP 420

QY 421 TPKKLPTTPPEKLAETTPPEKPAETTPPEELAPTTPEEPTTPPEEPAETTPKAAAPTPKE 480  
DB 620 TPKKLPTTPPEKLAETTPPEKPAETTPPEELAPTTPEEPTTPPEEPAETTPKAAAPTPKE 480  
QY 481 PAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 540  
DB 680 PAPTTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 540  
QY 541 EPNSTSDKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 600  
DB 740 EPNSTSDKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 600  
QY 601 KELAPTTGPTSTTSDKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 660  
DB 800 KELAPTTGPTSTTSDKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPKPAETTPK 660  
QY 661 PPTTKEPTTIHKSPEBSPBESAEPTPALENSPEKPGVPTTKPAATKPEMTTAKDKT 720  
DB 860 PPTTKEPTTIHKSPEBSPBESAEPTPALENSPEKPGVPTTKPAATKPEMTTAKDKT 720  
QY 721 TRDRLTPPETTAAPKMTKENTTEKTTESKITATTTQVSTTQDTTPPKITTLKT 780  
DB 920 TRDRLTPPETTAAPKMTKENTTEKTTESKITATTTQVSTTQDTTPPKITTLKT 780  
QY 781 TLAPRVTTTKTITTTTETIMNRPETAPKPKDRATNSKATTPPKORTKAPKPKTSKPKKT 840  
DB 980 TLAPRVTTTKTITTTTETIMNRPETAPKPKDRATNSKATTPPKORTKAPKPKTSKPKKT 840  
QY 841 MPRVAKRPTTPPKMTSTMBELNPTSRISIAAMLOTTRPQNTPSKLVENPKSEDAGG 900  
DB 1040 MPRVAKRPTTPPKMTSTMBELNPTSRISIAAMLOTTRPQNTPSKLVENPKSEDAGG 900  
QY 901 AEGEPHMLLPHYFMPVTPDMDYLPRVNOGIIINMLSDENINICGKRVYDGLTTURN 960  
DB 1100 AEGEPHMLLPHYFMPVTPDMDYLPRVNOGIIINMLSDENINICGKRVYDGLTTURN 960  
QY 961 GTLVAFRGHTFMLSPPSPSPARRITEWGIPIPIDVTFRCNCEGTFEFGKSOYWR 1020  
DB 1160 GTLVAFRGHTFMLSPPSPSPARRITEWGIPIPIDVTFRCNCEGTFEFGKSOYWR 1020  
QY 1021 TNDIKDGYPRIRKRGGLGQIYVAAALSTAKYNNPESYFEK 1064  
DB 1220 TNDIKDGYPRIRKRGGLGQIYVAAALSTAKYNNPESYFEK 1064  
RESULT 3  
AAB60568  
ID AAB60568 standard; Protein; 1404 AA.  
AC AAB60568;  
XX  
DT 21-APR-2001 (first entry)  
XX  
XX Human megakaryocyte stimulating factor (MSF, CACP).  
DE Human megakaryocyte stimulating factor (MSF, CACP).  
XX  
XX Human, CACP protein; campylocactyly-arthropathy-coxa vara-pericarditis;  
KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopachic;  
XX antiarthritic.  
OS  
XX Homo sapiens.  
XX  
XX  
PN WO200107068-A.  
XX  
PD 01-FEB-2001.  
XX  
XX 21-JUL-2000; 2000MO-US20002.  
XX  
PF 23-JUL-1999; 99US-0145328.  
PR 19-JUL-2000; 2000US-0145328.  
XX  
XX (UYCA-) UNIT CASE WESTERN RESERVE.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:15:40 ; Search time 107.17 Seconds  
(without alignments)  
953.822 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_25\_1404  
Perfect score: 7410  
Sequence: 1 QDSSCAGRCGEYSRDATC.....ABATRTSGOTLSKWTNCP 1380

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Genseq\_1101.\*  
1: /SIDSR/gcgdata/genseq/genseq/AA1980.DAT.\*  
2: /SIDSR/gcgdata/genseq/genseq/AA1981.DAT.\*  
3: /SIDSR/gcgdata/genseq/genseq/AA1982.DAT.\*  
4: /SIDSR/gcgdata/genseq/genseq/AA1983.DAT.\*  
5: /SIDSR/gcgdata/genseq/genseq/AA1984.DAT.\*  
6: /SIDSR/gcgdata/genseq/genseq/AA1985.DAT.\*  
7: /SIDSR/gcgdata/genseq/genseq/AA1986.DAT.\*  
8: /SIDSR/gcgdata/genseq/genseq/AA1987.DAT.\*  
9: /SIDSR/gcgdata/genseq/genseq/AA1988.DAT.\*  
10: /SIDSR/gcgdata/genseq/genseq/AA1989.DAT.\*  
11: /SIDSR/gcgdata/genseq/genseq/AA1990.DAT.\*  
12: /SIDSR/gcgdata/genseq/genseq/AA1991.DAT.\*  
13: /SIDSR/gcgdata/genseq/genseq/AA1992.DAT.\*  
14: /SIDSR/gcgdata/genseq/genseq/AA1993.DAT.\*  
15: /SIDSR/gcgdata/genseq/genseq/AA1994.DAT.\*  
16: /SIDSR/gcgdata/genseq/genseq/AA1995.DAT.\*  
17: /SIDSR/gcgdata/genseq/genseq/AA1996.DAT.\*  
18: /SIDSR/gcgdata/genseq/genseq/AA1997.DAT.\*  
19: /SIDSR/gcgdata/genseq/genseq/AA1998.DAT.\*  
20: /SIDSR/gcgdata/genseq/genseq/AA1999.DAT.\*  
21: /SIDSR/gcgdata/genseq/genseq/AA2000.DAT.\*  
22: /SIDSR/gcgdata/genseq/genseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7410	100.0	1404	13	MSF precursor. Sy
2	7410	100.0	1404	22	Human megakaryocyt
3	7410	100.0	1404	22	Human megakaryocyt
4	6834	92.2	1299	22	Human EST encoded
5	3484	47.0	902	22	Human MSF-derived
6	2081	28.1	452	16	Human megakaryocyt
7	1707.5	23.0	472	22	Bovine MSF ortholo
8	1188	16.0	5179	22	AA60566
9	981	13.2	763	21	AA624516
10	950	12.8	1664	19	AA38942
11	805	10.9	1325	22	AA43106
					Peptide #2327 enco
					Peptide #2327 enco

12	752	10.1	182	12	AAR10872
13	715	9.6	763	18	AAW31852
14	645.5	8.7	4412	21	AAV53666
15	613	8.3	111	13	AAR26050
16	612	8.3	572	18	AAW31855
17	555.5	7.5	844	7	AA60570
18	542	7.3	807	21	AAV54467
19	520.5	7.0	788	21	AAV54466
20	513	6.9	1837	21	AAW11726
21	506.5	6.8	744	9	AAW82875
22	498.5	6.7	1721	21	AAW11727
23	498	6.7	1721	19	AAW48299
24	496.5	6.7	2971	21	AAW41231
25	489.5	6.6	2972	22	AAW50363
26	489.5	6.6	3118	22	AAW50362
27	488.5	6.6	826	13	AAR26042
28	488	6.6	617	22	AAW16458
29	488	6.6	617	22	AAW04187
30	488	6.6	957	21	AAV59288
31	488	6.6	957	22	AAW24513
32	485	6.5	1127	22	AAW55541
33	476.5	6.4	652	9	AAW82874
34	467.5	6.3	511	22	AAW14883
35	467.5	6.3	511	22	AAW27312
36	467.5	6.3	511	22	AAW02607
37	454.5	6.1	1012	20	AAV17406
38	454.5	6.1	2819	22	AAW35408
39	450.5	6.1	378	12	AAW14160
40	448	6.0	1237	21	AAW81609
41	447.5	6.0	3266	21	AAW42491
42	446.5	6.0	378	12	AAW14162
43	446.5	6.0	750	20	AAW05477
44	446.5	6.0	2665	22	AAW14533
45	446.5	6.0	2665	22	AAW26950

## ALIGNMENTS

RESULT 1	
ID AAR26049	standard; Protein; 1404 AA.
XX	
AC AAR26049;	
XX	
DT 02-FEB-1993 (first entry)	
XX	
DE MSF precursor.	
XX	
KW Megakaryocyte colony stimulating factor; meg-CSF;	
KW stability; proteolytic cleavage; adhesion; alternative splicing.	
XX	
OS Synthetic.	
XX	
FH Key	Location/Qualifiers
FT Region	1..26
FT Region	/label= Exon_I
FT Region	26..67
FT Region	/label= Exon_II
FT Region	67..107
FT Region	/label= Exon_III
FT Region	107..157
FT Region	/label= Exon_IV
FT Region	157..200
FT Region	/label= Exon_V
FT Region	200..1141
FT Region	/label= Exon_VI
FT Region	1411..1166
FT Region	/label= Exon_VII
FT Region	1166..1212
FT Region	/label= Exon_VIII
FT Region	1213..1266
FT Region	/label= Exon_IX

Protein encoded by  
Mycobacterium tube  
Sequence g1/101742  
MSF-K130.. Synthet  
Mycobacterium tube  
Sequence of the Fa  
Aml1 acid sequenc  
Amino acid sequenc  
Cryptosporidium pa  
Bioadhesive precur  
Portion of Cryptos  
Cryptosporidium pa  
Human OREX ORF995  
Human SRCAP. Homo  
P. yoelii SSP2 ant  
Peptide #2892 enco  
Peptide #2869 enco  
Human MUC11 polype  
9900P predicted am  
Human protein sequ  
Bioadhesive precu  
Peptide #1317 enco  
Peptide #1349 enco  
Peptide #1289 enco  
Human atrophin-1 r  
Human 07C627 gene  
PRP 378. Trilicium  
Streptococcus pneu  
Human OREX ORF2355  
PRP encoded by clo  
C. albicans Rbt1 p  
Peptide #867 enco  
Peptide #867 enco



FT Region 1266..1331  
FT /label= Exon\_X  
FT Region 1331..1373  
FT /label= Exon\_XI  
FT Region 1373..1404  
FT /label= Exon\_XII  
PN MO9213075-A.  
XX 06-AUG-1992.  
PD 17-JAN-1992: 92MO-US00433.  
XX 18-JAN-1991: 91US-0643502.  
PR 10-SEP-1991: 91US-0757022.  
XX (GEMT ) GENETICS INST INC.  
PA  
XX Clark SC, Gesner TG, Hewick RM, Jacobs K, Turner K;  
PI WPI, 1992-284660/34.  
DR N-PDB; AAQ27223.  
XX  
XX New human mega-karyocyte stimulating factors - for treating  
PT immune deficiencies, cancer, exposure to radiation or drugs,  
PT bacterial and viral infections, etc.  
XX  
XX Claim 1, 2 and 3; Fig 1; 87pp; English.  
XX  
CC The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II,  
CC III and IV encodes megakaryocyte stimulating factor (MSF). This  
CC sequence is modified by the addition of an N-terminal sequence encoding  
CC a secretory leader, an initiating methionine preceding exon II and a  
CC terminating codon following exon IV. The cDNA sequence given contains  
CC sequences derived from human megakaryocyte colony stimulating factor  
CC (meg-CSF). Exon I contains the initiating methionine, and encodes a  
CC classical mammalian protein secretion signal sequence. The sequence  
CC encoding the original meg-CSF includes exons II-IV and is thought to  
CC terminate in the region between amino acid residues 134 - 147. The  
CC primary transcript of this gene may be cleaved in different ways to  
CC yield a family of mRNA's each encoding a different MSF protein. Exons  
CC V and VI are thought to be related to the activity of the factor and  
CC are also implicated in the stability, folding and processing of the  
CC molecule. These exons are also thought to play a role in the observed  
CC synergy of MSF with other cytokines. Exons V - XII are believed to be  
CC implicated in the processing or folding of the appropriate structure of  
CC the resulting factor, ie. one or more of these exons may contain  
CC sequences which direct proteolytic cleavage, adhesion, organisation of  
CC the cellular matrix or extracellular matrix processing. Both naturally  
CC occurring and non-naturally occurring MSF's may be characterised by  
CC various combinations of alternatively spliced exons from this sequence,  
CC with the exons spliced together in differing orders to form different  
CC members of the MSF family.  
XX  
XX  
SQ Sequence 1404 AA:  
  
Query Match 100.0%; Score 7410; DB 13; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ODLSSACGCGEGYSRATNCNDYNCQHYMCPCDFKVCYTAELSKCRCESEFERGEC 60  
DB 25 qdlsscgrcggyrldcncdyncqhyccpdkfvcvcaelsckrciesfergrec 84  
QY 61 DCDACCKRYDCPDYEFCAEVHNPTSPSSKRAPPSGASQRTKSTKSPSPNPKK 120  
DB 85 dcdacckrydcpcdyefcaeivhnptspsskrappsgsqclstkspspnnkk 144  
QY 121 TRKYIESEITREHVSNEOESSSSSSSTIMKIKSSKNSAANRELQKLVKDNK 180  
DB 145 tkryieeetleehvsneogesssssstlwkikssksnsaanelqklkvkdnk 204

QY 181 KNRKKKPKPPVNDVDEAGSLGNDGFKVTPTDSTQHNKYSTSPKTTAKPINRPSL 240  
DB 205 knrkkkpkppvndvdeagslgndgfkvtptdstqhnkystspkttakpinrpsl 264  
QY 241 PPSNDTSKETSLTVNKEITVEETKETTNNKQSTDGKERTTSARETOSIEKTSAKDLAPT 300  
DB 265 ppsndtsktsltvnkettvetkettlnkqstdgkertsaretsaktsakdlapt 324  
QY 301 SKVLAKPTKAEITTGAPALTTPKEPTPTTPKPEASTTKEPTTIKSAPTTPKEAPPT 360  
DB 325 skvlakptkaeittgapalttpkeptpttpkpeasttkepttiksapttpeappt 384  
QY 361 TTKSAPTPKEAPPTTKBPAPTTPKEAPPTTKBPAPTTPKAPPTTKBPAPTTPKRA 420  
DB 385 tksaptpkeappttkbpapttpkeappttkbpapttpkapattpkrappttkpa 444  
QY 421 PTPKPEAPPTTPKEAPPTTPKPEAPPTTPKPEAPPTTPKAPPTTPKAPPTTPK 480  
DB 445 ptpkpeappttpkeappttpkpeappttpkpeappttpkapattpkapattpk 504  
QY 481 EPAPTTKESPPTTPKEAPPTTKSAPTTTKSAPTTTKSAPTTTKESPPTTPKEAPT 540  
DB 505 epapttkesspttpkeappttksaptttksaptttksaptttkesspttpkeapt 564  
QY 541 PKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRA 600  
DB 565 pkrappttpkrapttpkrapttpkrapttpkrapttpkrapttpkrapttpckl 624  
QY 601 TPTTPKLAPTTPTEKAPPTTPPEELAPTTPEEPPTTPPEEPAPPTPKAANPTKEAPPT 660  
DB 625 tpttpklappttptekapttppeelapttpeeppttppeepapptpkaaaptkeappt 684  
QY 661 PKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKRAPPTTPKEPTST 720  
DB 685 pkrappttpkrapttpkrapttpkrapttpkrapttpkrapttpkrapttpkeptst 744  
QY 721 TSDKPAATTPKGAAPTPKRAPPTTPKRAPPTTPKGAAPTPKRAPPTTPKRAPKLAP 780  
DB 745 tsdkpaattpkgapttpkrapttpkrapttpkrapttpkgapttpkrapttpkrpklap 804  
QY 781 TTTKGPSTTSDBAPPTTPKETAATTPKETAATTPKPAETTPKPAETTPSEVSTPTTK 840  
DB 805 tttkgpsttsdbappttpketaattpketaattpkpaettpkpaettpsevspttk 864  
QY 841 EPTTIHKSPESTPELSAETTPKALNSPKRPGVPTTKTAATKPEATTAKDKTTERDL 900  
DB 865 epttihkspesstpeelsaettpkalnspkrgvpttktaatkpeattakdkttterdl 924  
QY 901 RTTPETTPAKPKMTKETATTEKTESKIRATTOVSTTODPTPKITLKTTLAPK 960  
DB 925 rtptettpakpkmtketattekteskiratovsttoddptpkitlkttlapk 984  
QY 961 VTTTKKTIITTEIMNKPEETAKPRDRATNSKATTPKPKOKTKAPKRTSTKPKTMPRVR 1020  
DB 985 vttkktittteimnkpeetakprdratnskattpkpkoktkapkrstskpkmprvr 1044  
QY 1021 KPKTTPPKKMTSTMPLEANTSRIAEAMLOTTTRPNQTPNSKLVENPKSEADGAGSET 1080  
DB 1045 kpkttpkkmtstmpleantsriaeamlotttrpnqtpnslkvevnpkseadagaset 1104  
QY 1081 PHMLLRPHVMPETVTPMOLVRPNQGIILNPLSDETINCNCKPVDGLTTLNGLTYA 1140  
DB 1105 phmlrrphvmpetvtpmolvrpnqgiilnplsdeitncnckpvdglttlnngltva 1164  
QY 1141 FRGHYFMMLSPSPSPSPARRITTEWAGIPSIDVFTVCNCEGKFFPKDSQYVHFTNDIK 1200  
DB 1165 frghyfmmlspspsparrittewagipsidvftvcncegkffkdsqyvfhftndik 1224  
QY 1201 DAGYPRKIFKGFGLGQIYAALSTAKYKMPESVYFFKRGSGSIQOYTYKQEPVQKCPGR 1260  
DB 1225 dagypkrfkfgfglgqiyaalstakympeesvyffkrgsgsiqoytykqepvqkcpgr 1284



QY 1261 RPALANPYGEMTOVRRRERERAIIGSPQTHTRIQYSPARLAYKQDKVLAHNEKYSIIIMR 1320  
DB 1285 RPALHPYVGYEMTQYRRIRFERAIGSPGCHTIRIIGSPARLAYGDKGVLAHNEKYSIIIMR 1344  
QY 1321 GLPNVNTSAISLPNIRKPKDGYDYAFSKNDQYINIVPSTARITRSGQTLSTKYMVNC 1380  
DB 1345 GJPNVNTSAISLPNIRKPKDGYDYAFSKNDQYINIVPSTARITRSGQTLSTKYMVNC 1404

## RESULT 2

AA060568  
ID AA060568 standard; Protein: 1404 AA.

AA060568;

27-APR-2001 (first entry)

Human megakaryocyte stimulating factor (MSF, CACP).

Human CACP protein; campodactyl-arthropathy-coxa vara-pericarditis;

MSF; megakaryocyte stimulating factor; synovial lubricant;

chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

antiarthritic.

Homo sapiens.

MO200107068-A1.

01-FEB-2001.

21-JUL-2000; 2000MO-US20002.

23-JUL-1999; 199US-0145328.

19-JUL-2000; 2000US-0145328.

(UTCA-) UNIV CASE WESTERN RESERVE.

Warman MI.

WPI: 2001-182721/18.

New composition comprising the campodactyl-arthropathy-coxa

vara-pericarditis protein in combination with an anesthetic, useful for

treating osteoarthritis or as lubricants of tissue and joints

Example 1: Page -: 34pp: English.

The invention relates to a method of treating osteoarthritis via the  
administration of a composition comprising the campodactyl-arthropathy-  
coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.  
The composition may further comprise a local anesthetic. The composition  
of the invention may be administered via intra-articular or intravenous  
injection. The human CACP protein is identified in the invention as  
being megakaryocyte stimulating factor (MSF). The gene encoding  
CACP protein (MSF) is located on chromosome 1q25-31, and mutations in  
this gene are responsible for the heritable disorder campodactyl-  
arthropathy-coxa vara-pericarditis, in which patients have synovial  
hyperplasia without evidence of inflammation. CACP protein (MSF)  
acts as a synovial lubricant, and can be used to lubricate tissue and  
joints in the treatment of osteoarthritis. The composition may be  
applied to reduce the symptoms of osteoarthritis (e.g., joint pain,  
loss of range of movement or joint damage). The present sequence  
represents human megakaryocyte stimulating factor (MSF, CACP protein).  
Note: This sequence is not given in its entirety in Figure 4 of the  
specification, although a GenBank accession number was given. This  
sequence was therefore obtained from GenBank (U00316).

Sequence 1404 AA:

Query Match 100.0%; Score 7410; DB 22; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY ODLSGAGRGEGYSRDATCNCOYNOHMECCPDFKRVCTAELSCGRPFESREREC 60  
DB 25 GDLSSCAGRCGEGYSRDATCNCOYNOHMECCPDFKRVCTAELSCGRPFESREREC 84  
QY 61 DCAOACKRYDKCCPDYEFSCAEVHNPTSPSSKAPPGASQTIKSTKRSPPPKK 120  
DB 85 ddaqckkykccpdysfaevhnptsppsaktkpppsasqtlksctkrspppk 144  
QY 121 TKVIESEETTEHSHVSENDESSSSSSSSSTIWKIKSSKNSAANRELQKLKVDNR 180  
DB 145 tkvleeseetleeshsvsenessssssssstlwkikssksaanrelqklkvdnr 204  
QY 181 KNRRTAKKRPKPPYVDAGSGLDGFKVTPPTSTQTHKVKYSTSPKTIKAKINRPSL 240  
DB 205 knrtakrkppkppvyvdagsgldgfkvtptstqthkvkstspktiakinp 264  
QY 241 PPNSDYKETSLSLVNKEETVETKETTNNKQSTDGKEKTSKETSQISAKDLAPT 300  
DB 265 ppsdtketslslvnketvettntnkqstdgketsketsketsakdlapt 324  
QY 301 SKYLAKEPKEAETTTGPAITTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPT 360  
DB 325 skylakepkeaeetttgpaaittpkepttpkepasttpkepttiiksapt 384  
QY 361 TTSAPPTPKKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAIT 420  
DB 385 ttsapptpkkappttpkepaaittpkepaaittpkepaaittpkepaait 444  
QY 421 PTPPKKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKE 480  
DB 445 ptppkkappttpkepaaittpkepaaittpkepaaittpkepaaittpke 504  
QY 481 EPAPTTPKEPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTP 540  
DB 505 epapttpkeppttpkepaaittpkepaaittpkepaaittpkepaaittpke 564  
QY 541 PKKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 600  
DB 565 pkkappttpkepaaittpkepaaittpkepaaittpkepaaittpkepa 624  
QY 601 TPTPKKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTP 660  
DB 625 tptpkkappttpkepaaittpkepaaittpkepaaittpkepaaittpke 684  
QY 661 PKKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 720  
DB 685 pkkappttpkepaaittpkepaaittpkepaaittpkepaaittpkepa 744  
QY 721 TSDKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKE 780  
DB 745 tsdkappttpkepaaittpkepaaittpkepaaittpkepaaittpkepa 804  
QY 781 TTTKGPSTTSKAPPTTPKEPAITTPKEPAITTPKEPAITTPKEPAITTPKEPA 840  
DB 805 tttkgpsttskappttpkepaaittpkepaaittpkepaaittpkepaait 864  
QY 841 EPTTIHKSDESTPELSAPPPKALENSPKSGVTTTPAATKREBMTTAAOKTTERDL 900  
DB 865 epttihkstpeelsapppkalenspksgvtttpaatkrebmmttaaktterdl 924  
QY 901 RTTPPTTTAAPKMTKETAATTEKTESKITATTTQVSTTODTTPFKITTKTTTLAPK 960  
DB 925 rtppptttaapkmteattekteskitatttqvsttoddtpfkittktttlap 984  
QY 961 VTTTKKTTTTEIIMKPEETAKPKDRATNSKATTPKOKPTKAPKPTSTKPKTTPRVR 1020  
DB 985 vttkktttteiimkpeetakpkdratnskattpkpkptkaptkptstckpk 1044  
QY 1021 KPKTTPPKKMTSTWPELNPSTRIAEAMLOTTPRNOHNSKLVYNNKSDAGABEET 1080  
DB 1045 kpktppkkmtstwpeelnpstriaeamlottpnrhnsklvynnksgagabebet 1104

